Jean-Pascal **Gayant**

AIDE-MÉMOIRE

Micro-économie

- Décisions du consommateur et du producteur
- Équilibre
- Monopole et oligopole



Tout le catalogue sur
www.dunod.com



ÉDITEUR DE SAVOIRS



5 rue Laromiguière, 75005 Paris
www.dunod.com

ISBN 978-2-10-071855-9

Table des matières

À Armelle, Anatole et Clémentin

Présentation

1 Qu'est-ce que la microéconomie ?

1.1 Les agents microéconomiques

La **microéconomie** est une discipline dont l'objet est d'étudier le comportement rationnel des agents économiques.

Les **agents économiques** sont des individus ou organisations qui prennent part à des relations économiques : production, consommation ou échange.

Les agents économiques sont donc :

- les **producteurs** de biens et services, c'est-à-dire les entreprises de toutes tailles (qui produisent des bâtiments, des automobiles, des meubles, des téléphones, des ordinateurs, des communications téléphoniques, des accès à internet, des transports, des services à domicile, etc.), les agriculteurs, éleveurs et pêcheurs (qui produisent des fruits, légumes, viandes et poissons), les médecins (qui produisent des diagnostics et des soins), les artistes, les sportifs professionnels (qui produisent des spectacles), les militaires, les policiers (qui produisent du maintien de l'ordre), les juges (qui produisent de la justice), les assureurs (qui produisent des couvertures assurancielles), les écrivains (qui produisent des textes), les journalistes (qui produisent de l'information), etc. et aussi, dans une certaine mesure, les partis politiques (qui produisent des projets d'organisation sociale et des candidats à la représentation des citoyens) et les institutions religieuses (qui produisent des dogmes, des rites et des cérémonies).

- **les consommateurs** des biens et services listés ci-dessus, c'est-à-dire chacun d'entre nous.

Il conviendrait d'ajouter une troisième catégorie d'agents économiques : la **puissance publique** (c'est-à-dire l'État et les collectivités territoriales). Précisément, la puissance publique est un agent supposé agir selon la volonté des citoyens-consommateurs et ne jamais poursuivre d'objectifs qu'elle se serait elle-même fixée. La réalité est plus complexe. La puissance publique, mise en place pour fournir des services que des marchés fonctionnant sur un mode traditionnel ne pourraient pas fournir (défense nationale, ordre public, justice, éclairage public, minima sociaux, etc.) ou des services dispensés au regard de l'importance du bénéfice commun qu'ils induisent (éducation), est une organisation qui parfois « perd de vue » sa raison d'être. Il en résulte une dérive où les administrations et agences publiques s'assignent elles-mêmes de nouveaux objectifs, cherchent à accroître leur taille pour les remplir et créent des mécanismes de contrôle de la satisfaction des objectifs en question (qui nécessitent de nouvelles ressources, elles-mêmes sujettes à la dérive décrite ci-dessus). La puissance publique est un agent dont l'objectif est parfois difficile à identifier ou préciser. En effet, l'objectif de la puissance publique doit être construit sur la base de l'agrégation des objectifs des consommateurs-citoyens. Une telle agrégation est complexe en théorie (par exemple, la solution du choix sur la base d'un vote majoritaire peut s'avérer très insatisfaisante) comme en pratique (les représentants élus ont un mandat « général » et non mandat pour une ou plusieurs décisions particulières ; de plus, ils poursuivent des objectifs personnels propres).

1.2 La méthode de la microéconomie

Les agents, producteurs et consommateurs, dont nous allons étudier le comportement, sont supposés rationnels. Définir la **rationalité** n'est pas une mince affaire. Est-ce qu'un pratiquant de vol en *wingsuit* (ce sport extrême où des individus frôlent, en chute libre, des parois rocheuses à des vitesses vertigineuses avant d'atterrir à l'aide d'un parachute) est un agent rationnel ? Est-ce qu'un individu avare et épargnant maladif qui, finalement, meurt sur un « tas d'or » sans laisser de descendance est un agent

rationnel ? Face à la grande diversité des comportements observés, est-il seulement possible de définir ce qu'est un comportement rationnel ?

Ce défi est celui de la science économique depuis ses origines, mais surtout depuis le XIXe siècle.

Dans le courant du XIXe siècle, est né l'**individualisme méthodologique**, c'est-à-dire l'approche selon laquelle chaque individu a des goûts, des préférences qui lui sont propres, et qu'en conséquence il prend des décisions qui peuvent différer de celles de son voisin. Auparavant, on raisonnait plus volontiers sur le comportement de classes : l'individu n'existait que comme élément d'un groupe dont le comportement était supposé homogène. Avec l'évolution des idées, inspirées par la philosophie des Lumières, il est apparu de plus en plus intenable de supposer que les individus puissent agir de manière non autonome, comme des automates préprogrammés au sein d'un groupe social.

L'individualisme méthodologique s'est donc imposé et avec lui l'**utilitarisme.** Par utilitarisme, on désigne l'approche en vertu de laquelle chaque agent cherche systématiquement à obtenir la plus forte utilité possible lorsqu'il décide des actions qu'il va entreprendre. Le mot « utilité » doit être entendu au sens large. Ce que chaque individu recherche, ce sont les actions qui maximisent sa satisfaction et/ou minimisent sa peine. Il espère accroître sa satisfaction en consommant des biens ou services marchands, certes, mais aussi en consacrant du temps à des activités non marchandes (promenade, jeu, etc.), et parfois mêmes délibérément altruistes (engagement associatif et/ou charitable, temps consacré à l'éducation de ses enfants, etc.). Pour l'économiste, il n'y a guère d'idée préconçue sur ce qui procure de la satisfaction aux différents individus ; en revanche le constat qui s'impose est celui d'une très grande variété d'aspirations et de sources de satisfactions. Le défi est donc de cerner un socle de « règles de comportement » que tous respectent en dépit de la grande variété des goûts et des préférences. C'est le point de départ de la modélisation des préférences individuelles.

La **modélisation**, où représentation simplifiée de la réalité, est une construction théorique reposant sur des hypothèses.

Dans le cas présent, les hypothèses sont des **axiomes de comportement** dont on postule qu'ils sont respectés par tous les individus. La liste de ces postulats permet au microéconomiste de délimiter les contours de ce qu'il entend par « rationalité individuelle ». Si, dans les faits, certains de ces axiomes sont transgressés par des individus, la question se pose de savoir s'il s'agit de comportements très rares, imputables à de manifestes erreurs de jugements ou si, au contraire, la très grande majorité des individus transgresse régulièrement ces axiomes. Dans le second cas, le modélisateur a vocation à s'interroger sur la pertinence du ou des axiomes incriminés et à envisager des aménagements et/ou modifications de la liste. Une fois que le modélisateur « tient » une liste satisfaisante d'axiomes de comportements, il peut s'attaquer à la traduction numérique des préférences : il s'agit d'établir les propriétés d'une fonction numérique qui représenterait les préférences de l'individu hypothétique dont on modélise le comportement. La fonction numérique en question a pour caractéristique d'associer une valeur numérique à chaque décision possible de l'agent, valeur numérique telle que la satisfaction éprouvée par l'agent sera d'autant plus forte que la valeur numérique sera elle-même élevée. En d'autres termes, on s'attache à construire une fonction de satisfaction encore appelée fonction d'utilité.

2 Les notions mathématiques utiles en économie

2.1 Fonction d'une variable

Le cas le plus simple de fonctions numériques est celui de fonctions d'une seule variable. Une fonction d'une seule variable est de la forme $y = f(x)$. Une telle expression décrit, pour toute valeur d'une variable $x \in \mathbb{R}$, quelle valeur de la variable $y \in \mathbb{R}$ lui est « associée » par la relation fonctionnelle f. On peut représenter graphiquement cette relation fonctionnelle dans un repère $(0 ; x ; y)$, en faisant figurer la variable x en abscisse et la variable y en ordonnée (voir figure 1).

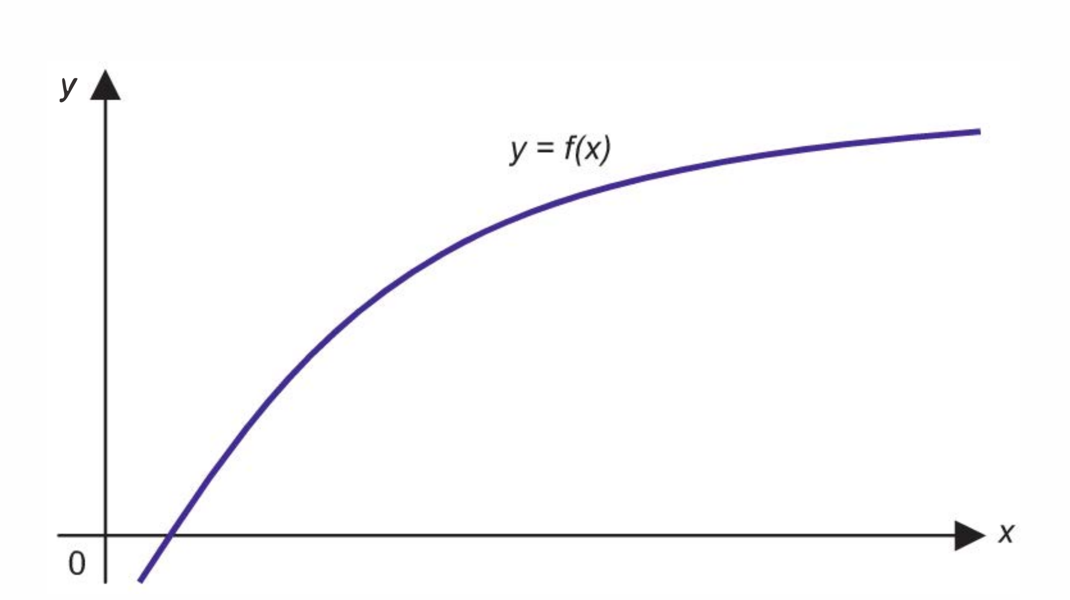


Figure 1 Représentation graphique d'une fonction d'une seule variable

2.2 Fonction de plusieurs variables

Dans certaines circonstances, il s'agit d'associer une valeur y à une collection ou « vecteur » de n valeurs $(x_1 ; x_2 ; \ldots ; x_n) \in \mathbb{R}^n$. On utilise alors une fonction de plusieurs variable de la forme : $y = f(x_1 ; x_2 ; \ldots ; x_n)$. On peut, dans le cas où $n = 2$, représenter graphiquement cette relation fonctionnelle dans un repère $(0 ; x_1 ; x_2 ; y)$, en faisant figurer la variable x_1 en abscisse, la variable x_2 en ordonnée, et la variable y en côte (voir figure 2).

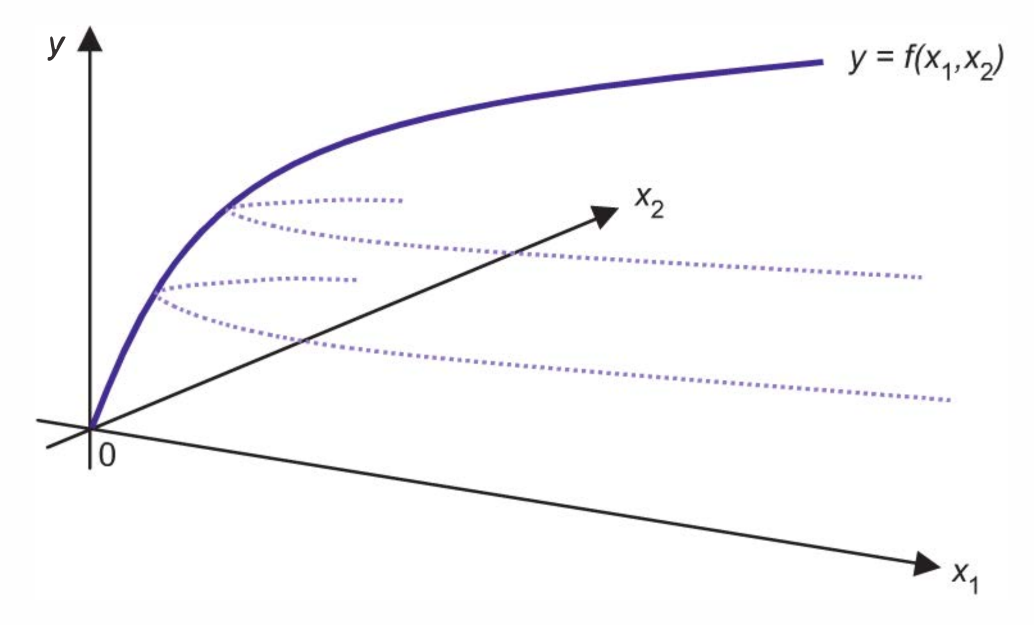


Figure 2 Représentation graphique d'une fonction de plusieurs variables

2.3 Dérivée et primitive d'une fonction d'une seule variable

La dérivée première d'une fonction d'une seule variable f en un point x_0 particulier est définie comme la limite du rapport de l'accroissement de cette fonction relativement à l'accroissement de sa variable lorsque ce dernier tend vers 0. La dérivée de f au point x_0 est notée $f'(x_0)$:

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

Graphiquement, la dérivée d'une fonction en un point est égale à la pente de la tangente à la courbe représentative de cette fonction en ce point (voir figure 3).

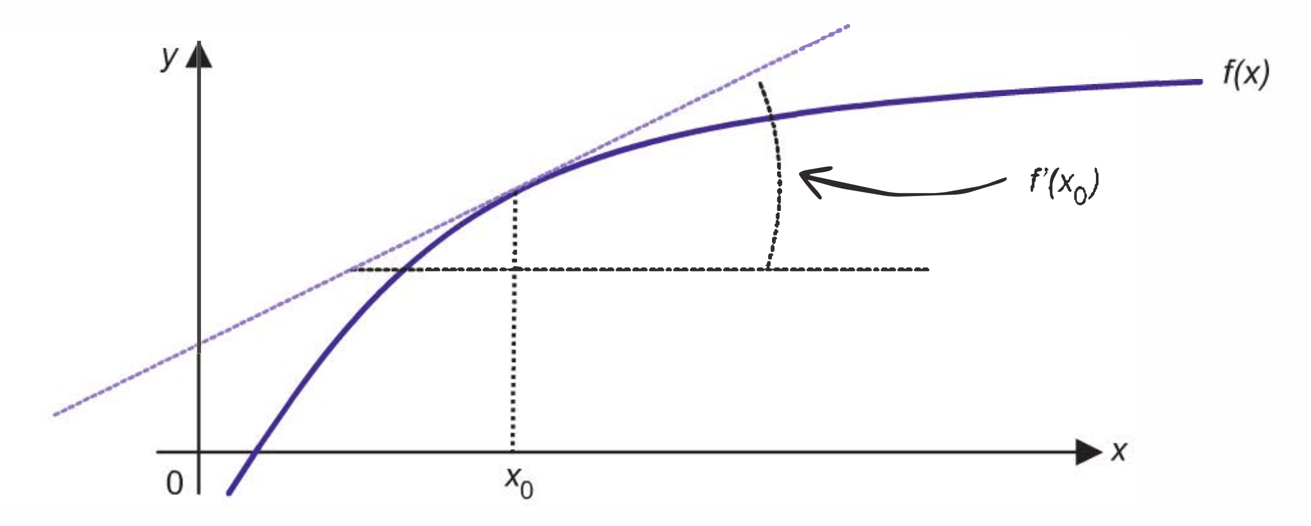


Figure 3 Dérivée d'une fonction d'une seule variable

Sur la figure 3, la fonction f est croissante. La tangente à la courbe en un point particulier sera donc une droite elle-même croissante. La dérivée première de la fonction en ce point, égale à la pente de la tangente, est donc positive.

Réciproquement, la primitive d'une fonction est la forme fonctionnelle dont la dérivée première est la fonction considérée (à une constante K près).

Ainsi, la primitive de la fonction f' est la fonction f :

$$f(x_0) = \int_{-\infty}^{x_0} f'(x)dx$$

Graphiquement, la primitive d'une fonction entre deux points est égale à l'aire sous la courbe représentative de cette fonction entre ces deux points. Sur la figure 4, la surface mauve est égale à $\int_{x_0}^{x_0'} g(x)dx$.

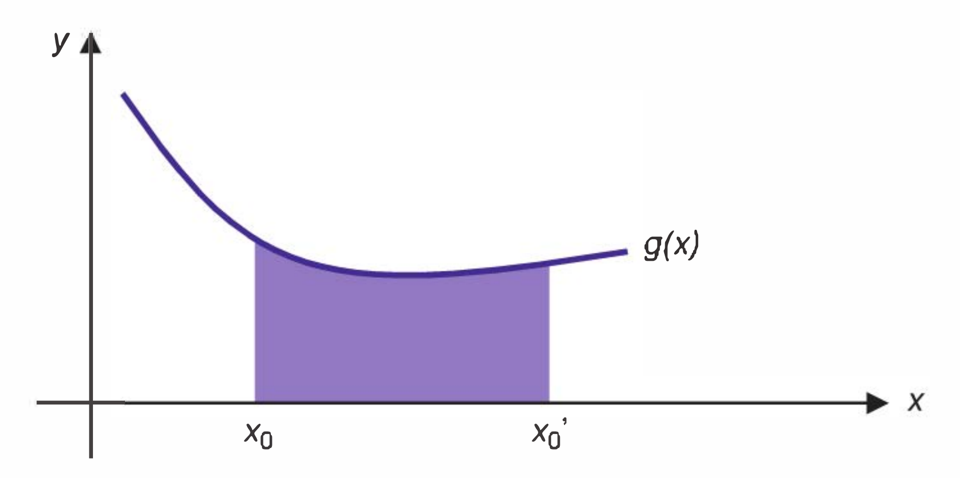


Figure 4 Primitive d'une fonction (zone mauve)

2.4 Dérivées partielles et différentielle d'une fonction de plusieurs variables

Soit $y = f(x_1 ; x_2 ; ... ; x_n)$ une fonction de n variables réelles. La dérivée partielle de la fonction f relativement à la variable x_i, notée $\frac{\partial f}{\partial x_i}$, est définie comme la limite du rapport de l'accroissement de la fonction relativement à l'accroissement de la variable x_i lorsque ce dernier tend vers 0. Ainsi :

$$\frac{\partial f}{\partial x_i} = \lim_{\Delta x_i \to 0} \frac{f(x_1;x_2;...;x_i + \Delta x_i;...;x_n) - f(x_1;x_2;...;x_i;...;x_n)}{\Delta x_i}$$

Afin de pouvoir manipuler facilement les dérivées partielles (qui caractérisent le comportement de la fonction f au voisinage de tout « point » dans les n dimensions qui permettent de l'identifier), on construit un outil qui décompose la variation infinitésimale de la variable y en la contribution des différents accroissements infinitésimaux des variables x_1, x_2, ..., x_{n-1} et x_n.

Si on note dx_i, l'accroissement infinitésimal de la variable x_i, la variation infinitésimale de la variable y est :

$$dy = df(x_1 ; x_2 ; ... ; x_n) = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + ... + \frac{\partial f}{\partial x_n} dx_n$$

L'expression $df(x_1 ; x_2 ; ... ; x_n)$ est la différentielle de la fonction f.

2.5 Concavité, convexité d'une fonction

Une fonction f est **concave** si l'image, par la fonction f, d'une combinaison linéaire convexe d'abscisses est supérieure à la combinaison linéaire convexe des images de ces abscisses.

Ainsi, dans le cas d'une fonction d'une seule variable, f est concave si :

$$\forall\, x, x' \in \mathbb{R} \text{ et } \forall\, \lambda \in [0 ; 1], f(\lambda x + (1-\lambda) x') \geq \lambda f(x) + (1-\lambda) f(x')$$

Sur la figure 5 est représentée une fonction concave.

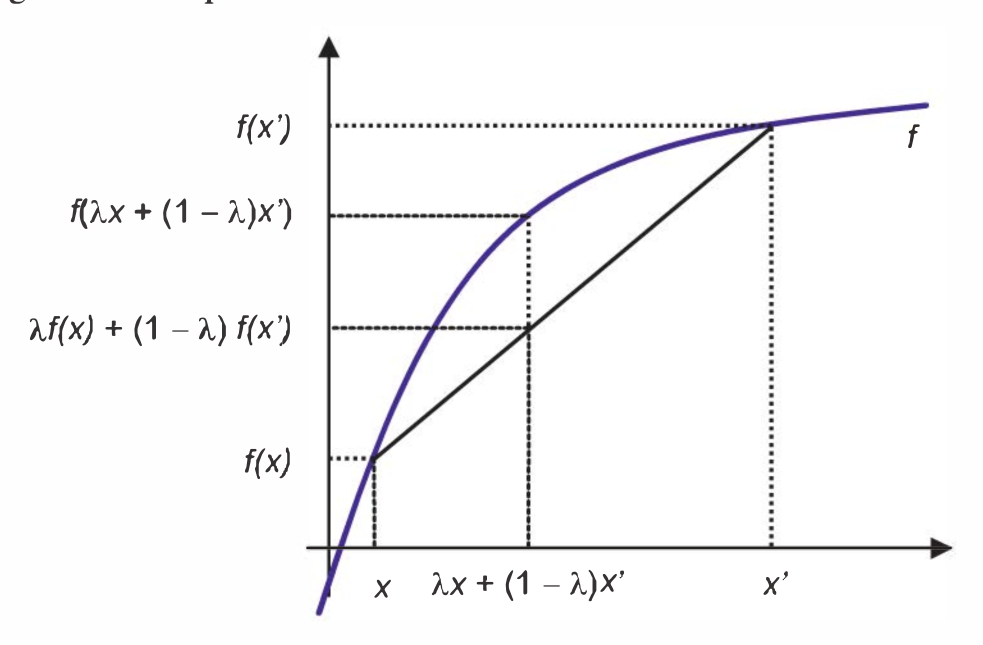


Figure 5 Fonction concave

Théorème : Une fonction d'une seule variable f est concave si $f''(x) \leq 0$.

Dans le cas d'une fonction de plusieurs variables, la définition de la concavité est formellement identique, à ceci près que les abscisses x et x' sont désormais des vecteurs de $\mathbb{R}^n$.

Dans le cas d'une fonction d'une seule variable, f est **convexe** si :

$$\forall\, x, x' \in \mathbb{R} \text{ et } \forall\, \lambda \in [0\,;1], f(\lambda\, x + (1-\lambda)\, x') \leq \lambda f(x) + (1-\lambda)\, f(x')$$

Théorème : Une fonction d'une seule variable f est convexe si $f''(x) \geq 0$.

Enfin, de la même manière, dans le cas d'une fonction de plusieurs variables, on peut écrire la définition de la convexité en remplaçant les scalaires x et x' par des vecteurs de $\mathbb{R}^n$.

2.6 Convexité d'un ensemble

Un ensemble est convexe si toute combinaison linéaire convexe d'éléments de l'ensemble appartient également à l'ensemble. Ainsi, un ensemble E est convexe si :

$$\forall\, x, x' \in E \text{ et } \forall\, \lambda \in [0\,;1], \lambda x + (1-\lambda)\, x' \in E$$

Sur la figure 6 apparaissent un ensemble E convexe et un ensemble F non convexe.

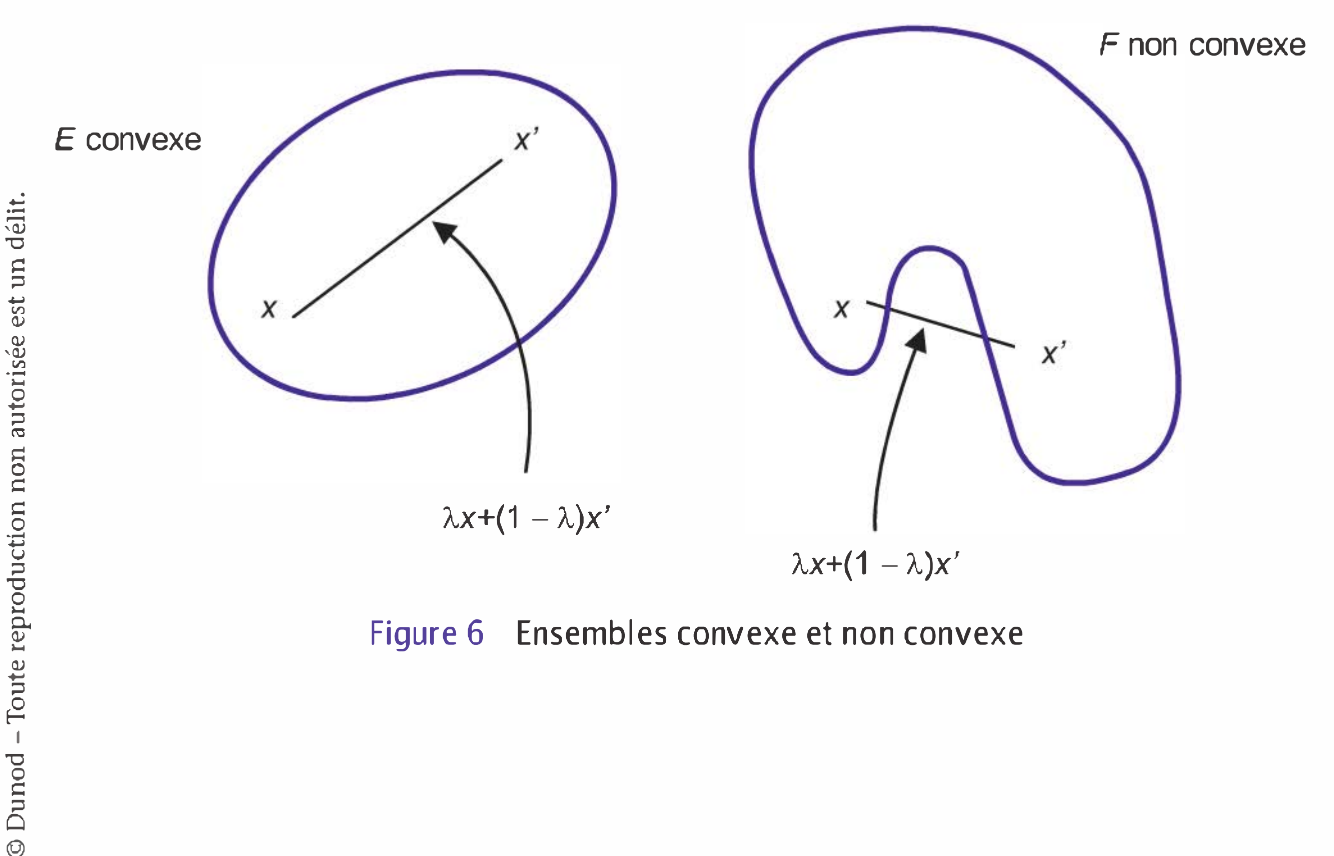


Figure 6 Ensembles convexe et non convexe

2.7 Maximum, minimum d'une fonction

Commençons par le cas de fonctions d'une seule variable.

Une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ admet un **maximum** en x^* si $\forall\, x \in \mathbb{R}, f(x^*) \geq f(x)$.

Théorème (Conditions nécessaires d'optimalité) : Si f admet un maximum en x^*, alors nécessairement :

- $f'(x^*) = 0$ condition du 1er ordre ;
- $f''(x^*) \leq 0$ condition du 2e ordre.

Les conditions ci-dessus sont des conditions nécessaires. Elles ne sont pas suffisantes pour qu'un point x^* qui les vérifierait soit le maximum absolu de la fonction f. Ceci peut se comprendre en observant la figure 7 (le maximum absolu de la fonction f est en effet atteint au point x^{**}).

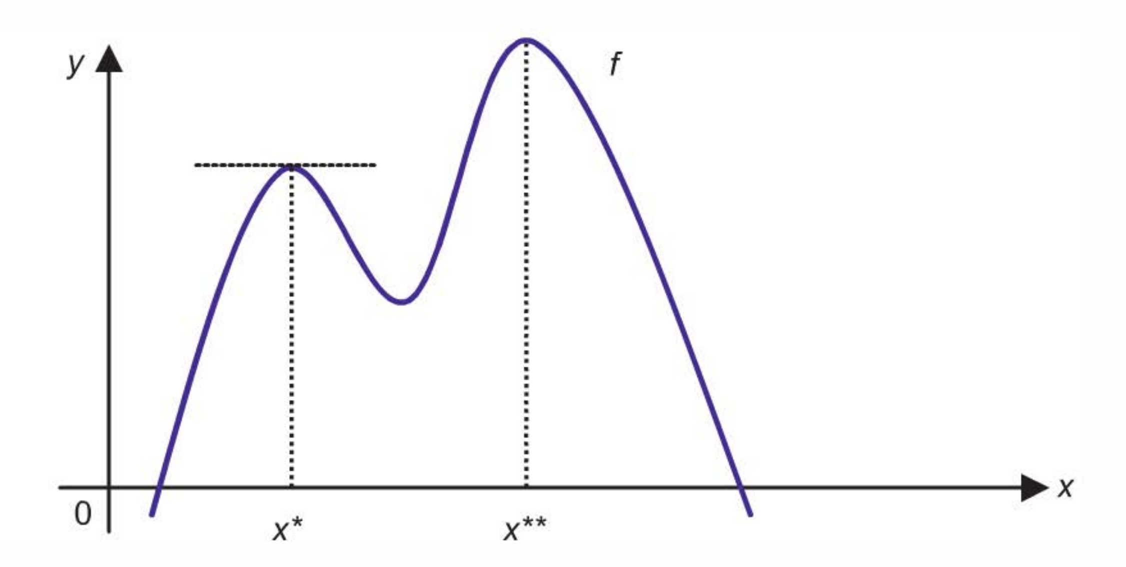


Figure 7 Le point x^* vérifie les conditions nécessaires d'optimalité mais la fonction n'atteint pas son maximum en ce point

Théorème (Conditions nécessaires et suffisantes d'optimalité) : Si f est concave alors x^* est un maximum $\Leftrightarrow f'(x^*) = 0$ (condition de 1er ordre) et $f''(x^*) < 0$ (condition de 2e ordre).

Sur la figure 8 apparaît le maximum d'une telle fonction f concave pour laquelle la dérivée première de la fonction au point x^* est nulle et la dérivée seconde en ce point est négative.

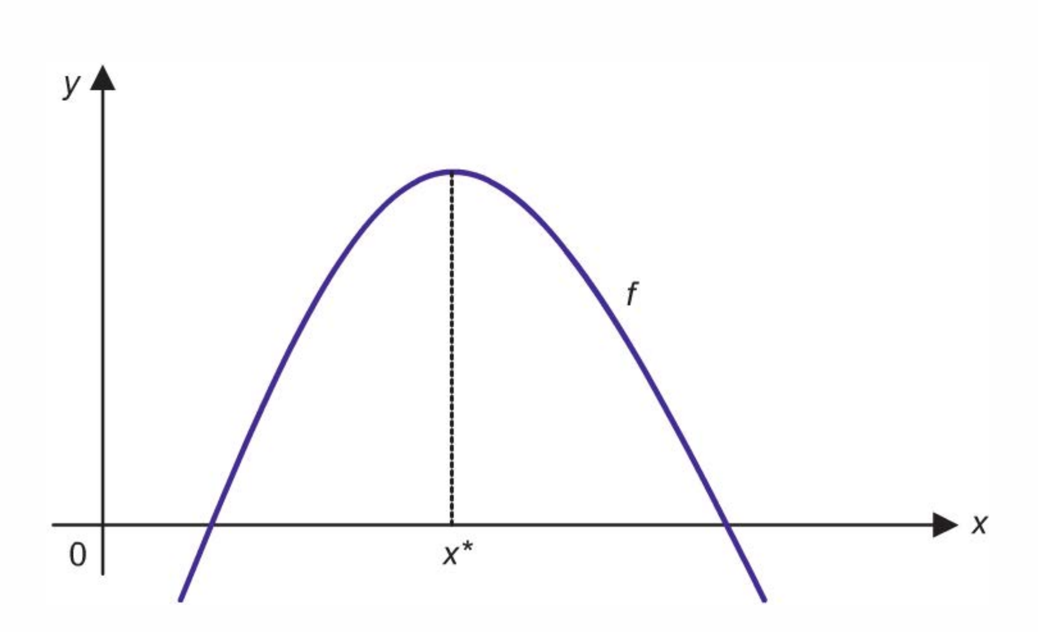


Figure 8 Conditions nécessaires et suffisantes d'optimalité

De la même manière, on peut définir et caractériser le minimum d'une fonction d'une seule variable.

Une fonction $f: \mathbb{R} \to \mathbb{R}$ admet un **minimum** en x^* si $\forall\, x \in \mathbb{R}, f(x^*) \leq f(x)$.

Théorème (Conditions nécessaires d'optimalité) : Si f admet un minimum en x^*, alors nécessairement :

- $f'(x^*) = 0$ condition du 1er ordre ;
- $f''(x^*) \geq 0$ condition du 2e ordre.

Théorème (Conditions nécessaires et suffisantes d'optimalité) : Si f est convexe alors x^* est un minimum $\Leftrightarrow f'(x^*) = 0$ (condition de 1er ordre) et $f''(x^*) > 0$ (condition de 2e ordre).

Enfin, il est possible de définir le maximum ou le minimum d'une fonction de plusieurs variables en remplaçant le scalaire $x \in \mathbb{R}$, par le vecteur $x \in \mathbb{R}^n$. Une fonction $f: \mathbb{R}^n \to \mathbb{R}$ admet un maximum (respectivement un minimum) en x^* si $\forall\, x \in \mathbb{R}^n, f(x^*) \geq f(x)$ (respectivement $f(x^*) \leq f(x)$).

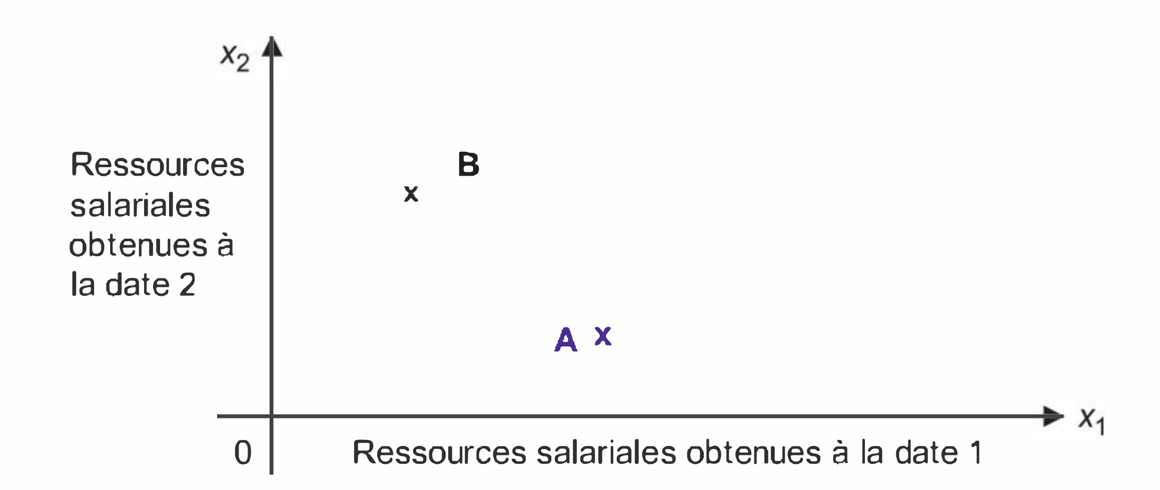
Quant aux conditions de 1er et de 2e ordre apparaissant dans les théorèmes de conditions nécessaires ou de conditions nécessaires et suffisantes d'optimalité, elles sont, dans l'esprit, analogues mais nécessitent d'être déclinées dans un contexte à n dimensions. On remplace ainsi la dérivée



première de la fonction par le vecteur des dérivées (partielles) premières au point considéré (le « gradient ») et la dérivée seconde par la matrice des dérivés (partielles) secondes au point considéré (la matrice « hessienne »). Il s'agit alors d'examiner la nullité du gradient et le caractère (semi-)défini négatif (respectivement positif) de la matrice hessienne.

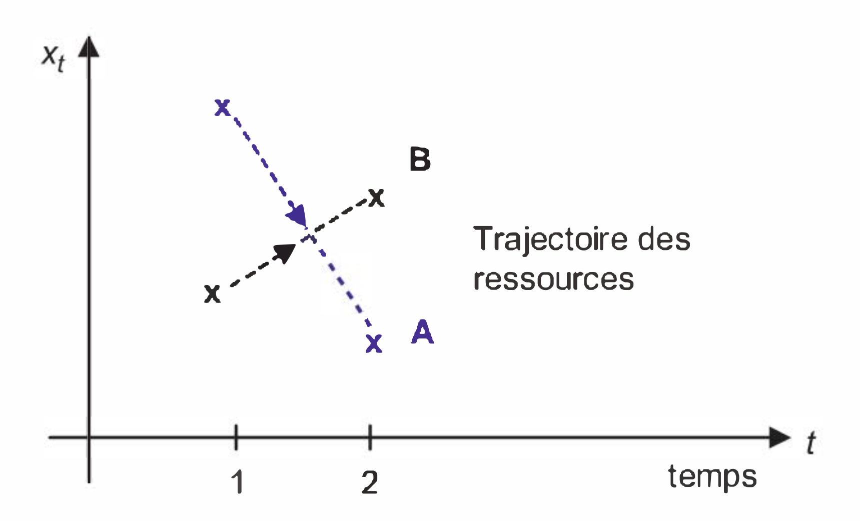
Comment « lire » un graphique

Pour bien comprendre ce qu'un graphique représente, il faut être attentif aux indications qui figurent en abscisse et en ordonnée. Dans certains cas, les points qui apparaissent dans le repère dessiné sont d'interprétation difficile ou abstraite. Prenons, par exemple, le cas où figure, en abscisse, la valeur prise par une variable x à une 1re date et où figure, en ordonnée, la valeur prise par cette même variable x à une 2e date, postérieure à la 1re. Pour faciliter la présentation, on désigne par x_1 la valeur prise par la variable en 1re date et par x_2 la valeur prise par la variable à la 2e. À titre d'illustration, considérons que x_1 mesure les ressources salariales obtenues dans des jobs étudiants lors de la première année d'études supérieures et que x_2 mesure les ressources salariales obtenues lors de la seconde année d'études. N'importe quel point dans le repère $(0 ; x_1 ; x_2)$ représente une **trajectoire** salariale, c'est-à-dire la chronique des ressources obtenues lors des deux années d'étude.



Imaginons que les points A et B représentent les trajectoires salariales de 2 étudiants, nommés Anatole et Bérénice. Ce que l'on observe sur le

graphique est qu'Anatole a obtenu d'importantes ressources salariales en 1re période et de beaucoup plus faibles en 2e période. À l'inverse, Bérénice a obtenu de plus fortes ressources en 2e période qu'en 1re. Son profil de ressources est, cependant, plus régulier. Pour mieux le comprendre, traçons les trajectoires de ressources des deux étudiants dans un repère avec le temps en abscisse et les ressources en ordonnée.



En microéconomie, les graphiques sont souvent utilisés pour représenter des situations abstraites. Un autre exemple frappant est celui des biens contingents. Dans un repère peuvent ainsi figurer les quantités consommées d'un même bien dans des états de la nature différents, c'est-à-dire pour des circonstances différentes dans un avenir proche (selon la réalisation ou non d'un événement particulier, de manière aléatoire). Un certain effort est nécessaire pour convenablement appréhender ce qu'illustrent certaines représentations graphiques. Il convient de toujours s'interroger sur le sens de ce qui, sous nos yeux, paraît parfois trop élémentaire. Enfin, il ne faut pas confondre un déplacement *le long d'une courbe* ou le déplacement *d'une courbe*. Dans le chapitre 8, à propos de l'équilibre d'un marché, nous aurons l'opportunité de pleinement distinguer les deux types de mouvements.

3 Ce que vous trouverez dans ce livre

Dans un 1er chapitre, nous étudierons la manière dont il est possible de formaliser les préférences d'un consommateur rationnel et de traduire numériquement ces préférences à l'aide d'une fonction d'utilité. Dans les chapitres suivants (2 à 4), nous analyserons le comportement du consommateur rationnel : sujet à une contrainte budgétaire (ses ressources ne sont pas illimitées), comment le consommateur construit-il sa décision optimale de consommation ? Une fois les principes généraux de ce type de décisions étudiés (chapitre 2), nous en étudierons les déclinaisons à l'arbitrage consommation-loisir, à l'arbitrage intertemporel de consommation et à la décision en présence d'incertitude (chapitre 3). Dans un 4e chapitre, nous envisagerons la genèse de la demande globale pour un bien ou service, obtenue par agrégation des demandes individuelles. Le chapitre 5 sera consacré à l'équilibre général d'une économie d'échange : quelles sont les motivations qui poussent deux agents à échanger un bien contre un autre ? Dans le chapitre 6, nous nous tournerons vers le comportement de l'entreprise : nous analyserons comment une firme organise la production d'un bien ou service. Dans le chapitre 7, nous étudierons la manière dont une entreprise construit sa décision d'offre en contexte concurrentiel, puis nous caractériserons, dans le chapitre 8, la notion d'équilibre d'un marché. Les chapitres 9 et 10 sont consacrés au fonctionnement de marchés non concurrentiels. Dans le chapitre 9, on envisage le cas d'un monopole, configuration dans laquelle une firme unique assure la production du bien ou service considéré tandis que dans le chapitre 10, on étudie des secteurs dans lesquels un petit nombre de firmes de tailles conséquentes rivalisent pour obtenir les parts de marché. Le chapitre 11 sera consacré à l'étude des biens publics et des effets externes dont la présence rend nécessaire l'intervention de l'État. Enfin, dans le chapitre 12, nous évoquerons les limites, imputables aux asymétries d'information, des constructions présentées jusque-là.

1 Les préférences et l'utilité

Mots-clés

Relation de préférence, fonction d'utilité, taux marginal de substitution, courbe d'indifférence, convexité des préférences.

Notre premier objectif est d'analyser la manière dont un consommateur, quel qu'il soit, prend ses décisions. Les biens et services consommés sont extrêmement variés et il en va de même pour les goûts des consommateurs. Il faut donc construire un cadre d'analyse où les consommateurs sont supposés respecter un noyau très restreint de règles de comportement, suffisamment large pour appréhender les préférences très variées des différents consommateurs. Il faut, dans un second temps, rendre cette construction utilisable pour traiter de questionnements économiques : comment un consommateur décide-t-il de dépenser son revenu entre les différents biens de consommation ? Que se passe-t-il si le prix de l'un des biens augmente ou si ses ressources diminuent ? Peut-il exister des biens dont la consommation augmente quand le prix s'accroît ?… Pour rendre le traitement de ces questions possible, nous traduisons numériquement les préférences du consommateur, c'est-à-dire que nous caractérisons sa fonction d'utilité, de satisfaction.

1 La relation de préférence et les axiomes de comportement

1.1 Les préférences

La première étape consiste donc à supposer que chaque individu a, lorsqu'il achète des biens, des préférences qui lui sont propres. Ces préférences

subjectives peuvent être traduites par une « **relation de préférence** » qui porte sur différents biens et groupes de biens (paniers). C'est un outil mathématique régi par certaines règles et dont on suppose qu'il représente tous les éléments de la subjectivité, des goûts de l'individu considéré. Certaines règles concernant ces préférences sont supposées communes à tous. Le choix de ces règles délimite les contours de ce qu'il convient, ou non, de considérer comme rationnel. Une modification, même infime, de la liste ou de la teneur des règles postulées par le modélisateur peut déplacer sensiblement la limite entre les comportements admis comme rationnels et ceux réputés irrationnels. C'est pourquoi il faut définir avec grand soin les propriétés supposées être respectées par l'ensemble des agents économiques. Une fois définies ces propriétés, la seconde étape de notre travail consiste à résumer, à l'aide d'une fonction numérique, la hiérarchie qu'établit la relation de préférence de l'individu : c'est ce que nous désignons par **la traduction numérique des préférences**.

Comment traduire numériquement les préférences d'un individu exprimées sur des « paniers de biens » ? Un panier de biens est une collection de biens ou services qu'un individu s'apprête à « consommer ». Le cas le plus simple est celui d'un panier de deux biens ; un bien étant noté x. Prenons l'exemple de jouets d'enfants. Un (jeune) individu apprécie les autos miniatures et les soldats de plomb. Nous désignerons par x_1 la quantité d'auto miniatures et par x_2 la quantité de soldats de plomb qu'il « consomme ». Un panier particulier sera donc noté $(x_1 ; x_2)$. Notre modèle s'attache à construire une relation de préférence qui permettra de décrire le choix de l'individu entre deux paniers uniquement constitués d'autos miniatures et de soldats de plomb, par exemple le choix entre les paniers $A = (5 ; 3)$ et $B = (2 ; 7)$. Le symbole utilisé pour matérialiser la relation de préférence est le signe $\succsim$. Si, aux yeux de l'individu le panier A est préférable au panier B, on écrira $A \succsim B$; si, à l'inverse, il préfère le panier B au panier A, on écrira $B \succsim A$. Pour être plus précis, le symbole ici utilisé désigne une relation de **préférence ou indifférence**. Cela signifie que lorsque l'on écrit que $A \succsim B$, on exprime que l'individu préfère le panier A au panier B ou que l'individu est parfaitement indifférent entre les deux paniers.

On appelle **relation de préférence (ou indifférence)** définie sur l'ensemble des paniers de n biens, la relation binaire qui incarne l'expression des préférences d'un consommateur entre toute paire de paniers de biens. Si, aux yeux du consommateur, le panier de biens A est préféré ou indifférent au panier de biens B, on note alors $A \gtrsim B$.

1.2 Les axiomes de comportement

Quelles sont les règles que tout individu rationnel est supposé respecter ? Le microéconomiste va postuler que tous les individus respectent quatre axiomes : la réflexivité, la transitivité, la complétude et la non-saturation. Présentons, une à une, ces règles de comportement.

- **Réflexivité** : N'importe quel panier de bien A est préféré ou indifférent à lui-même.
- **Transitivité** : Quels que soient 3 paniers de biens A, B et C, si A est préféré ou indifférent à B et si B est préféré ou indifférent à C alors nécessairement A sera préféré ou indifférent à C.
- **Complétude** : Quels que soient 2 paniers de biens A et B, soit A est préféré ou indifférent à B, soit B est préféré ou indifférent à A, soit les deux.
- **Non-saturation** : Si A et B sont deux paniers de biens tels que la quantité de chacun des biens disponibles est au moins aussi grande dans le panier A qu'elle ne l'est dans le panier B et s'il existe au moins un bien pour lequel la quantité de ce bien est strictement supérieure dans le panier A à ce qu'elle est dans le panier B, alors A est strictement préféré à B. On résume parfois cet axiome de non-saturation en disant que le consommateur « souhaite toujours obtenir des unités additionnelles de chacun des biens ou services » ou, plus laconiquement encore, qu'il « préfère obtenir *plus* plutôt que *moins* ».

Ces règles sont raisonnables et intuitives. Les propriétés de transitivité et de non-saturation sont essentielles. Elles sont néanmoins fragiles au sens où, dans certaines circonstances extrêmes, les individus rationnels sont susceptibles de les transgresser. Il en va ainsi, par exemple, de la propriété de non-saturation. Supposons que l'un des biens du panier soit une

mignardise très sucrée. La première mignardise sera très appréciée. La seconde également. La troisième pièce commencera à écœurer celui qui les déguste. Il ne parviendra à peine à finir la quatrième et refusera la cinquième, dégoûté. Ainsi, il existe des circonstances dans lesquelles la consommation d'une unité supplémentaire de bien n'apportera aucune satisfaction supplémentaire et, au contraire, la fera décroître. Néanmoins, nous allons considérer que, dans des circonstances « normales » (et dans le cadre canonique de la recherche du meilleur panier de biens sous un budget limité), l'axiome de non-saturation est pertinent.

Pour aller plus loin

Formulation mathématique des axiomes du consommateur

Il est possible d'écrire les 4 propriétés ci-dessus sous forme mathématique. Les spécifications des propriétés sont présentées, ci-dessous, dans le cas de paniers de n biens ($n \geq 2$).

- **Réflexivité**

$\forall A \in \mathbb{R}^n, A \succsim A$

- **Transitivité**

$\forall A, B, C \in \mathbb{R}^n, A \succsim B$ et $B \succsim C \Rightarrow A \succsim C$

- **Complétude**

$\forall A, B \in \mathbb{R}^n, A \succsim B$ ou $B \succsim A$

- **Non-saturation**

$\forall A = (x_1 ; x_2 ; \ldots ; x_n)$ et $B = (y_1 ; y_2 ; \ldots ; y_n) \in \mathbb{R}^n$ tels que $\forall i \in \{1, 2, \ldots, n\}$ $x_i \geq y_i$ et $\exists j \in \{1, 2, \ldots, n\}$ tel que $x_j > y_j$, alors $A \succ B$.

Remarque : La relation « $\succ$ » apparaissant dans l'axiome ci-dessus est la relation de préférence au sens strict. Elle se déduit de la relation de préférence ou indifférence par l'expression suivante : $\forall A, B \in \mathbb{R}^n, A \succ B \Leftrightarrow$ Non « $B \succsim A$ ».

2 La traduction numérique des préférences : la fonction d'utilité

Dès lors que les propriétés de réflexivité, transitivité, complétude et non-saturation sont satisfaites, il est possible de traduire numériquement les préférences de l'individu considéré, c'est-à-dire de caractériser une fonction numérique qui associe, à chaque panier de biens, une valeur d'autant plus grande que le panier est apprécié. Désignons par u cette fonction. Sa caractéristique fondamentale est que, si un panier A est préféré ou indifférent à un panier B, la valeur numérique $u(A)$ associée au panier A sera supérieure ou égale à la valeur numérique $u(B)$ associée au panier B. En termes mathématiques :

$$\forall\ A, B \in \mathbb{R}^n, A \gtrsim B \Leftrightarrow u(A) \geq u(B)$$

Chaque individu consommateur est donc caractérisé par une certaine fonction u. Puisque cette fonction u accorde au panier de biens ou services une valeur d'autant plus grande que la satisfaction éprouvée par le consommateur est forte, il semble assez naturel de la nommer fonction de satisfaction ou **fonction d'utilité**.

On appelle **fonction d'utilité** définie sur l'ensemble des paniers de n biens, la fonction numérique qui associe à tout panier de biens une valeur d'autant plus forte que le panier est apprécié par le consommateur considéré. Si $x = (x_1 ; x_2 ; \ldots ; x_n)$ désigne un panier de n biens, l'utilité éprouvée par le consommateur est notée $u(x_1 ; x_2 ; \ldots ; x_n)$.

Ce que mesure cette fonction d'utilité ne s'inscrit pas sur une échelle *cardinale* d'intensité. En d'autres termes, le niveau de satisfaction que mesure cette fonction n'est en rien comparable à la température que mesure un thermomètre ou au nombre de kilomètres parcourus que mesure le compteur d'une voiture. Il n'existe, en effet, aucune échelle universelle de la satisfaction, au contraire de ce dont on dispose pour mesurer les températures ou les distances.

L'une des raisons de l'inexistence d'une échelle de satisfaction est qu'il est impossible de s'accorder, entre deux individus, sur la hiérarchie entre les

« plaisirs » (et les peines) : pour l'un, le panier *A* sera préféré au panier *B*, pour l'autre le panier *B* sera préféré au panier *A*. Personne n'a raison ni tort, chacun à des goûts propres. Dès lors, il est vain d'espérer construire une unité de mesure de la satisfaction et la fonction d'utilité n'est donc pas un concept cardinal. La fonction d'utilité est en revanche un concept *ordinal* : elle indique des valeurs numériques dont l'ordre est parfaitement fidèle à l'ordre des préférences qu'exprime le consommateur entre les paniers de biens ou services. Reprenons l'exemple de notre petit consommateur qui exprime ses préférences sur des paniers uniquement constitués d'autos miniatures et de soldats de plomb et supposons que ce dernier préfère le panier $A = (5\ ;\ 3)$ au panier $B = (2\ ;\ 7)$. Peu importe que nous obtenions $u(A) = 1\ 000$ et $u(B) = 850$ ou $u(A) = 0{,}3$ et $u(B) = 0{,}25$, ce qui importe est que $u(A)$ soit supérieur à $u(B)$. Seule la hiérarchie entre les valeurs numériques a une signification ; les valeurs numériques prises isolément n'en ont aucune. D'un point de vue mathématique, la fonction d'utilité est d'ailleurs définie à une transformation croissante près, ce qui signifie que n'importe quelle fonction croissante transformant la fonction u permettra, elle aussi, de traduire correctement la hiérarchie des préférences du consommateur.

3 À quoi ressemble une fonction d'utilité ?

Les fonctions d'utilité peuvent prendre de multiples formes. Il y a néanmoins une exigence essentielle qui doit être satisfaite : la satisfaction éprouvée doit croître lorsque l'on augmente la quantité consommée d'un des biens ou services, *toutes choses égales par ailleurs* (c'est-à-dire, ici, lorsque les quantités consommées des autres biens ou services sont inchangées). En mathématiques, on dira que la fonction est croissante relativement à chacune de ses variables.

Prenons un exemple dans le cas de paniers de deux biens (les autos miniatures et les soldats de plomb) : $u(x_1\ ;\ x_2) = x_1^2 \times x_2$

Cette fonction est bien croissante relativement à chacune de ses variables. Si nous mesurons l'utilité associée au panier *A*, nous obtenons $u(A) = 5^2 \times 3 = 25 \times 3 = 75$. Si nous calculons l'utilité d'un panier qui

contiendrait 1 soldat de plomb supplémentaire (un panier A' composé de 5 autos miniatures et de 4 soldats de plomb), nous trouvons $u(A') = 5^2 \times 4 = 25 \times 4 = 100$. L'utilité a bien augmenté suite à l'augmentation de la quantité « consommée » de l'un des biens. Il en aurait été de même si nous avions augmenté la quantité consommée de l'autre bien.

Mathématiquement, une fonction de plusieurs variables sera croissante relativement à l'une de ses variables si sa dérivée partielle, relativement à cette variable, est positive :

Si $u(x_1 ; x_2 ; ... ; x_n)$ désigne une fonction de n variables réelles, u est croissante relativement à la variable x_i si $\frac{\partial u(x_1; x_2; ...; x_n)}{\partial x_i} \geq 0$.

Le fait que la fonction d'utilité soit croissante relativement à chacune de ses variables est une conséquence directe de l'axiome de non-saturation. Il est essentiel, pour traduire le fait que chaque consommateur préfère toujours disposer de quantités additionnelles de tous les biens, que l'utilité augmente lorsque la quantité consommée de n'importe quel bien s'accroît.

4 L'utilité marginale

L'**utilité marginale** d'un bien est le surcroît de satisfaction engendré par la consommation d'une unité supplémentaire de ce bien. On notera Um_i l'utilité marginale du bien i.

On peut en donner une définition plus précise : l'utilité marginale d'un bien est le surcroît de satisfaction engendré par la consommation d'une *quantité infinitésimale supplémentaire* de ce bien. Cette nuance nous révèle comment écrire, d'un point de vue mathématique, l'utilité marginale : en effet, la notion ci-dessus désignée correspond à la mesure de la façon dont la satisfaction croît quand on augmente, de manière infinitésimale, la quantité consommée d'un des biens. Or ceci est précisément la dérivée partielle de la fonction d'utilité. Ainsi, l'utilité marginale du bien i est $Um_i = \frac{\partial u(x_1; x_2; ...; x_n)}{\partial x_i}$.

La notion d'utilité marginale est essentielle pour comprendre la formation des décisions des individus : nous verrons dans le chapitre suivant que, pour constituer son panier de biens, le consommateur va allouer ses ressources en s'efforçant de sélectionner les biens et services au regard de leurs utilités marginales respectives. Mais l'utilité marginale d'un bien est (comme la fonction d'utilité) une notion ordinale. Ainsi, le surcroît de satisfaction mesuré en calculant l'utilité marginale d'un bien n'a pas de signification en soi. Ce n'est qu'en le comparant au surcroît de satisfaction engendré par la consommation d'une unité supplémentaire d'un autre bien que nous obtenons une information pertinente. C'est ce que rend possible le calcul du taux marginal de substitution entre deux biens.

5 Le taux marginal de substitution (*TmS*)

Le taux marginal de substitution entre deux biens est une mesure des proportions dans lesquelles un consommateur est prêt à échanger un bien contre un autre sans que ne soit modifié le niveau de sa satisfaction. On notera $TmS_{i/j}$ le taux marginal de substitution entre les biens *i* et *j*.

L'idée d'un taux de substitution entre deux biens est très simple. Reprenons l'exemple des autos miniatures et des soldats de plomb et supposons que l'enfant consomme le panier $B = (2 ; 7)$. En consommant le panier *B*, et en rappelant que $u(x_1 ; x_2) = x_1{}^2 \times x_2$, le jeune consommateur éprouve une satisfaction égale à $2^2 \times 7 = 4 \times 7 = 28$. Ce que nous cherchons à déterminer est la quantité de soldats de plomb qu'il faudrait donner à l'enfant pour compenser la renonciation à la jouissance de l'une des autos miniatures ; l'opération devant conduire à ce que le niveau d'utilité soit inchangé, c'est-à-dire qu'elle demeure égale à 28. En d'autres termes, si on lui retire une unité de bien 1, combien d'unités de bien 2 faut-il lui donner pour que sa satisfaction reste identique ? Si l'on note Δx_1 la variation du nombre d'unités de bien 1 consommé ($\Delta x_1 = -1$), quel accroissement Δx_2 du nombre d'unités de bien 2 permettrait de garder inchangée l'utilité ? Spontanément, on songe à construire un rapport de variations absolues : taux de

substitution ½ = $TS_{½} = \frac{\varnothing x_2}{\varnothing x_1}$ (ce quotient est nécessairement négatif puisque Δx_2 doit être positif si Δx_1 est négatif et vice versa). Dans l'exemple ci-dessus, il est facile de calculer la réponse car nous connaissons précisément l'équation de la fonction d'utilité : si l'on retire 1 auto miniature à l'enfant, il ne disposera plus que d'1 seule auto, ou, en d'autres termes $x_1 = 1$. Puisque $x_1{}^2$ est aussi égale à 1, la valeur de x_2, qui maintient égale à 28 le niveau d'utilité éprouvé est donc $x_2 = 28$. Il faut donc $\Delta x_2 = +21$ pour maintenir l'utilité à un niveau inchangé (il faudra accorder à l'enfant, qui dispose initialement de 7 soldats de plombs, 21 soldats supplémentaires puisque 7 + 21 = 28). Cela conduit à un taux de substitution $TS_{½} = \frac{+21}{-1}$ $= -21$.

L'outil que nous cherchons à définir dans cette section est un peu différent de ce qui a été présenté ci-dessus. Nous voulons, en réalité, construire un taux **marginal** de substitution entre les biens 1 et 2, c'est-à-dire calculer le nombre d'unités de bien 2 qu'il faut accorder au consommateur pour qu'il accepte de renoncer à la consommation d'une quantité **infinitésimale** de bien 1 (et, implicitement, que sa satisfaction reste inchangée). D'un point de vue mathématique, un taux marginal de substitution s'exprimera comme $TmS_{½} = \frac{dx_2}{dx_1}$. Le terme dx_i désigne une variation infinitésimale de la quantité consommée de bien i. Il est possible de montrer (cf. p. 37) que ce quotient $\frac{dx_2}{dx_1}$ est égal à $-\frac{Um_1}{Um_2}$, c'est-à-dire (au signe près) au rapport des utilités marginales des biens 1 et 2.

Propriété

Le taux marginal de substitution entre deux biens est égal (au signe près) au rapport des utilités marginales des biens. Mathématiquement, $TmS_{½} = -\frac{Um_i}{Um_j}$.

L'une des raisons qui conduit à privilégier un raisonnement sur des variations infinitésimales est la nécessité de disposer d'un outil qui permet de traiter le cas de biens divisibles. Les biens non divisibles (ou non fractionnables) sont les produits dont il n'est possible de consommer que des quantités « entières » (c'est-à-dire des nombres entiers) : automobiles, bouteilles de vin, vêtements, smartphones, trajets en train, accès internet, etc. On les oppose aux biens divisibles, dont il est possible de consommer des quantités « réelles » (c'est-à-dire des nombres réels, éléments de l'ensemble $\mathbb{R}$) : litres d'essence, kilos de tomates, mètres cubes de gaz, durées de communications téléphoniques, etc. Plus fondamentalement, le raisonnement sur des variations infinitésimales nous permettra d'exprimer, en toutes circonstances, les caractéristiques du comportement optimal du consommateur.

6 La représentation graphique des préférences

Pour donner une représentation graphique de ces préférences, nous allons comme précédemment, considérer les préférences des consommateurs lorsqu'elles s'expriment sur un panier de seulement deux biens (x_1 et x_2 désignant toujours les quantités consommées des deux biens). Pour représenter les préférences du consommateur dans un espace à seulement deux dimensions, nous allons devoir opter pour le concept de courbes d'iso-utilité, à l'image des courbes de niveau apparaissant sur les cartes topographiques ou des courbes isobares apparaissant sur les cartes de prévision météorologiques. Des courbes d'iso-utilité sont, par définition, des courbes dont tous les points représentent des paniers de biens apportant *un même niveau de satisfaction* au consommateur. Rappelons que les lignes de niveaux sur une carte représentent les positions situées *à une même altitude* (relativement au niveau de la mer) et que les courbes isobares représentent les lieux soumis *au même niveau de pression atmosphérique*. Les courbes d'iso-utilité, puisqu'elles regroupent l'ensemble des paniers de biens entre lesquels le consommateur est indifférent, sont aussi appelées **courbes d'indifférence**.

Une **courbe d'indifférence** est un ensemble de paniers de biens dont la consommation apporte un même niveau d'utilité à un consommateur particulier. Si l'on désigne par *K* ce niveau de satisfaction, l'équation de la courbe d'indifférence est $u(x_1 ; x_2) = K$.

Reprenons le cas de notre jeune consommateur qui joue avec des autos miniatures et des soldats de plomb. Situons, dans un plan où la quantité consommée de bien 1 (les autos) figure en abscisse et où la quantité consommée de bien 2 (les soldats) figure en ordonnée, l'ensemble des paniers lui apportant un même niveau de satisfaction *K* (figure 1). Dans le cas de biens indivisibles (figure 1), nous voyons apparaître une collection de points qui constitue l'ébauche d'une courbe **décroissante**. En effet, si l'on diminue le nombre de petits soldats mis à sa disposition, il faudra augmenter le nombre d'autos miniatures laissées à son libre usage pour que son niveau de satisfaction reste inchangé (et vice versa). Si les biens 1 et 2 sont des biens divisibles (figure 2), un tel ensemble d'iso-utilité va prendre la forme d'une courbe continue.

En tout point de la courbe apparaissant sur la figure 2, l'utilité éprouvée par le consommateur est égale à *K*.

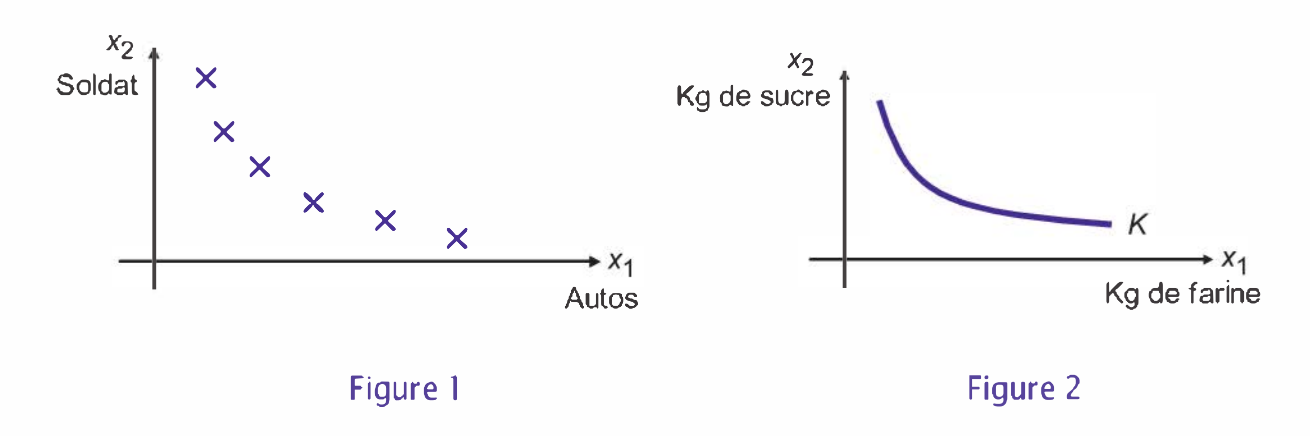


Figure 1 Figure 2

Dans le cas de paniers de 3 biens, il ne sera pas possible de raisonner à partir de courbes d'indifférences mais à partir de *surfaces d'indifférence* tracées dans un repère à 3 dimensions. On peut tenter de suggérer la forme de telles surfaces (figure 3) mais il ne sera guère commode de recourir à de

telles représentations graphiques pour décrypter la manière dont le consommateur prend ses décisions.

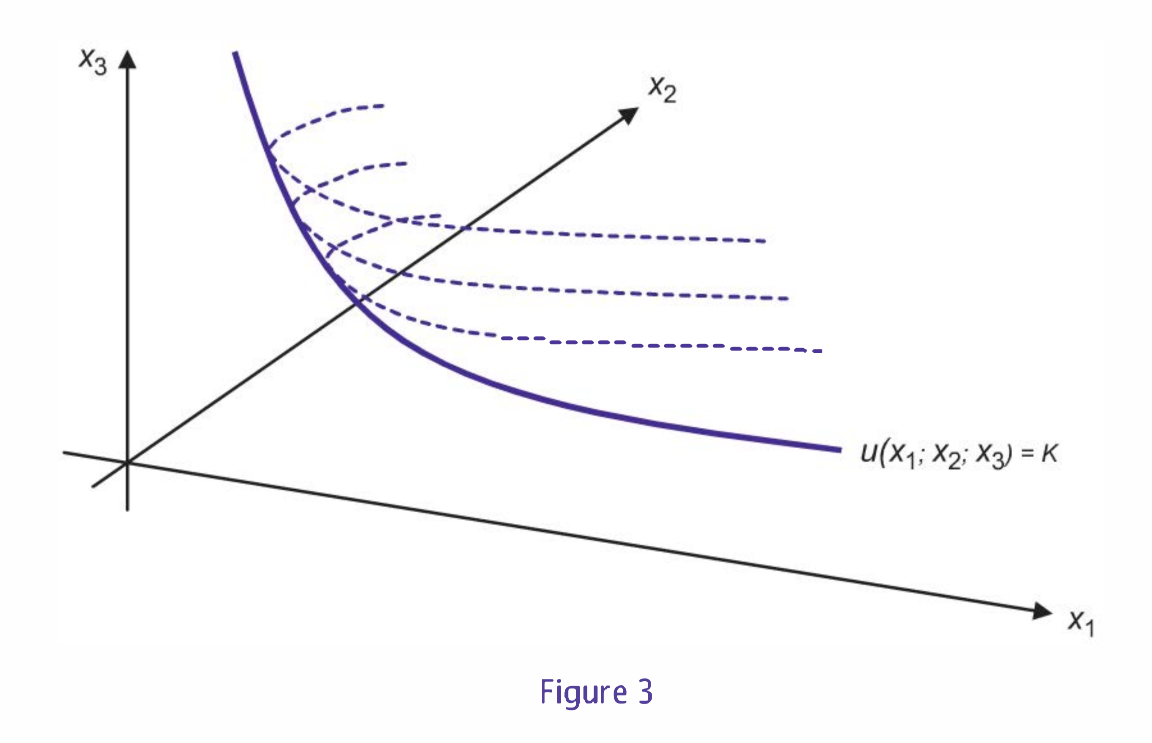


Figure 3

Mentionnons enfin que dans le cas de paniers de 4, 5, 6, … biens, il ne pourra plus être question que de représentations graphiques abstraites des ensembles d'iso-utilité apparaissant dans des repères à 4, 5, 6, … dimensions. Fort heureusement, nous pouvons amplement caractériser les traits fondamentaux de la décision du consommateur en raisonnant sur des paniers constitués de deux biens seulement. Pour cela, nous devons mieux connaître les propriétés des courbes d'indifférence.

6.1 Propriétés caractéristiques des courbes d'indifférence

Puisque nous supposons que tout consommateur respecte les axiomes de réflexivité, transitivité, complétude et non-saturation, nous pouvons affirmer que toutes les courbes d'indifférence possèdent trois propriétés principales.

Propriétés

Les courbes d'indifférence sont décroissantes, ne se croisent pas et correspondent à des niveaux d'utilité d'autant plus élevés que l'on se situe plus haut, vers la droite.

Les courbes d'indifférence sont décroissantes

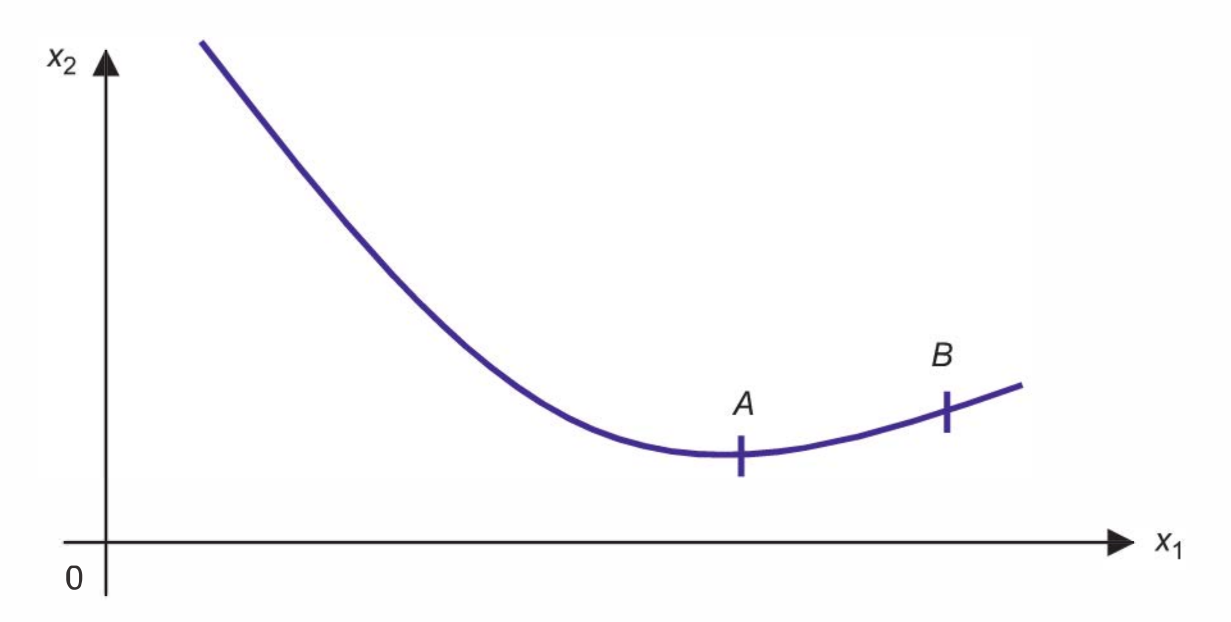


Figure 4 Ceci ne peut pas se produire

Pour montrer que ce qui apparaît sur le graphique ci-dessus ne peut pas se produire, on raisonne *par l'absurde*. Cela consiste à partir d'une « hypothèse absurde » qui affirme que les courbes d'indifférence auraient, comme sur le graphique, une partie croissante. En vertu d'une telle hypothèse, il existerait nécessairement des paniers A et B appartenant à une même courbe tels que, à la fois $x_1^B > x_1^A$, et $x_2^B > x_2^A$. Or, d'après l'axiome de non saturation, dans de telles circonstances, $B \succ A$ (le panier B serait strictement préféré au panier A), ce qui contredit l'hypothèse qu'ils appartiennent à la même courbe d'indifférence. L'hypothèse absurde étant contredite, il est dès lors établi que les courbes d'indifférence sont décroissantes.

Les courbes d'indifférence ne se croisent pas

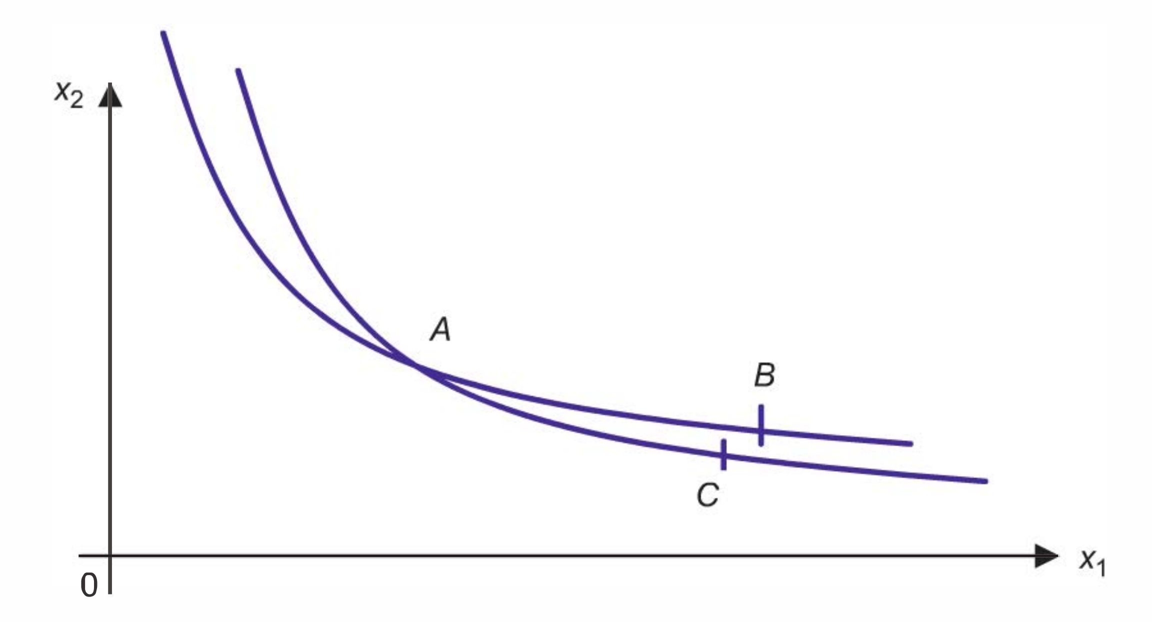


Figure 5 Ceci ne peut pas se produire

Comme précédemment, on peut raisonner par l'absurde pour montrer que des courbes d'indifférence distinctes (c'est-à-dire correspondant à des niveaux de satisfaction différents) ne se croisent. Ici, l'hypothèse absurde est que deux courbes d'indifférence distinctes se croisent en un point *A*. En vertu d'une telle hypothèse, il existerait nécessairement un panier *B* appartenant à la première courbe et un panier *C* appartenant à la seconde courbe tels que ceux représentés sur la figure 5. Puisque *A* et *B* appartiennent à une même courbe d'indifférence, nécessairement $A \sim B$ (le décideur serait indifférent entre les paniers *A* et *B*)[1] ; de même, *A* et *C* appartiennent à une même courbe d'indifférence donc $A \sim C$. D'après l'axiome de transitivité, nécessairement $B \sim C$, ce qui contredit l'hypothèse qu'ils appartiennent à des courbes d'indifférence distinctes. L'hypothèse absurde est donc là aussi contredite : des courbes d'indifférence distinctes ne croiseront jamais.

1. La relation « ~ » est la relation binaire d'indifférence. Elle se déduit de la relation de préférence ou indifférence par l'expression suivante : $\forall A, B \in \mathbb{R}^n$, $A \sim B \Leftrightarrow A \succsim B$ et $B \succsim A$.

■ Le niveau d'utilité est d'autant plus élevé que les courbes d'indifférence sont situées plus haut, vers la droite

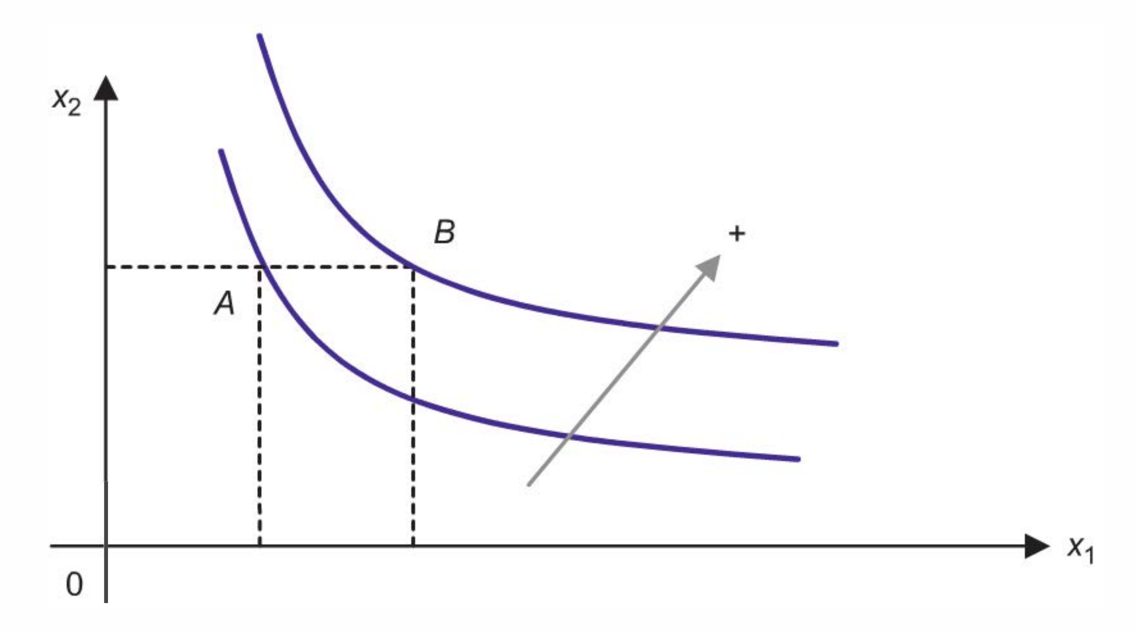


Figure 6

Considérons deux courbes d'indifférence appartenant à un même consommateur et identifions des paniers A et B tels que sur le graphique ci-dessus (B étant situé sur une courbe d'indifférence plus haute et plus à droite que celle sur laquelle est positionné A). Nous observons que les paniers A et B sont tels que à la fois $x_1^B > x_1^A$ et $x_2^B = x_2^A$. D'après l'axiome de non saturation $B \succ A$, ou ce qui équivalent $u(B) > u(A)$. Donc le niveau d'utilité est d'autant plus élevé que les courbes d'indifférence sont situées plus haut, vers la droite.

7 « Lire » le *TmS* sur les courbes d'indifférence

L'équation de toute courbe d'indifférence est $u(x_1 ; x_2) = K$. Cette écriture n'est pas très simple à manipuler. Il est plus habituel de travailler avec une fonction prenant la forme $x_2 = f(x_1)$ où x_2 désigne la variable située en ordonnée et x_1 la variable située en abscisse dans un repère $(0 ; x_1 ; x_2)$.

Il n'est, en général, pas difficile de transformer l'équation de la courbe d'indifférence pour obtenir une forme fonctionnelle standard. Dans l'exemple que nous utilisons, où $u(x_1 ; x_2) = x_1^2 \times x_2$, écrire $u(x_1 ; x_2) = K$, ou encore $x_1^2 \times x_2 = K$, est équivalent à écrire $x_2 = \frac{K}{x_1^2}$, ce qui correspond à une expression de type $x_2 = f(x_1)$ qu'il est facile de représenter graphiquement. La représentation graphique d'une telle fonction est exactement du type de celles que nous avons fait apparaître ci-dessus : une courbe continue, décroissante et strictement convexe. Nous allons aborder la question de la convexité de la fonction (et donc des préférences) dans la prochaine section. Ce sur quoi nous allons mettre l'accent ici est la possibilité de visualiser le *TmS* en un panier de biens quelconque en observant attentivement la courbe représentative de la fonction $x_2 = f(x_1)$ (ou, en d'autres termes, de la courbe d'indifférence $u(x_1 ; x_2) = K$). Nous allons nous appuyer sur une propriété mathématique qui énonce que, pour une fonction numérique représentée graphiquement dans un repère avec la variable x_1 en abscisse et la variable x_2 en ordonnée, le quotient $\frac{dx_2}{dx_1}$ mesure, en un point quelconque, *la pente de la tangente à la courbe en ce point*. Or, par construction, le quotient $\frac{dx_2}{dx_1}$ n'est autre que le taux marginal de substitution entre les biens 1 et 2, noté $TmS_{1/2}$.

Propriétés

Dans un repère avec la quantité consommée de bien 1 en abscisse et la quantité consommée de bien 2 en ordonnée, la pente de la tangente à la courbe d'indifférence en un point (un panier de biens) quelconque est égale au taux marginal de substitution entre les biens 1 et 2.

Cette propriété est représentée sur la figure 7.

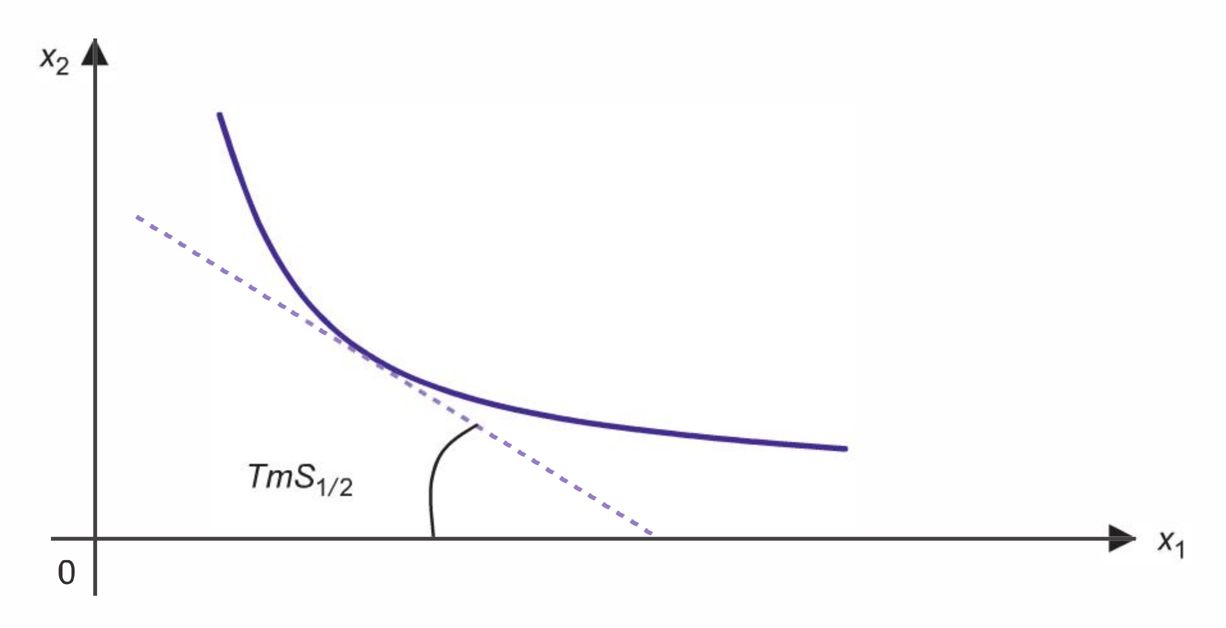


Figure 7

8 La convexité des préférences

8.1 Définition

Toutes les courbes d'indifférence que nous avons tracées jusqu'à présent possèdent le point commun d'être **convexes** (et même strictement convexes). Ceci traduit une *convexité des préférences* du consommateur. Cette hypothèse de convexité des préférences n'est pas universelle en microéconomie : les préférences de certains consommateurs exprimées sur certains types de paniers de biens peuvent ne pas être convexes. Des cas de préférences non strictement convexes existent (étudiées dans la deuxième partie de cette section), soit parce que le type de biens composant le panier ne se prête pas à l'existence de cette propriété, soit parce que les goûts du consommateur s'écartent de ce qui s'apparente à un comportement majoritaire (sans qu'on puisse qualifier le consommateur d'« irrationnel »).

Les préférences d'un décideur sont (strictement) **convexes** si pour tout panier de biens *A*, l'ensemble des paniers préférés ou indifférent à *A* est (strictement) convexe.

Pour comprendre cette définition, reprenons le cas de paniers de deux biens. Considérons un panier A quelconque dans le plan $(0\ ;\ x_1\ ;\ x_2)$. Essayons de faire apparaître l'ensemble des paniers préférés ou indifférent à *A*. Il apparaît très vite que cet ensemble, qui se situe plutôt en haut et plutôt à droite du point *A*, sera précisément délimité par la courbe d'indifférence passant par *A*. Si l'on trace cette courbe, il apparaît que tous les points situés plus haut et plus à droite que la courbe d'indifférence (courbe incluse) constituent l'ensemble des paniers préférés ou indifférent à *A* (figure 8).

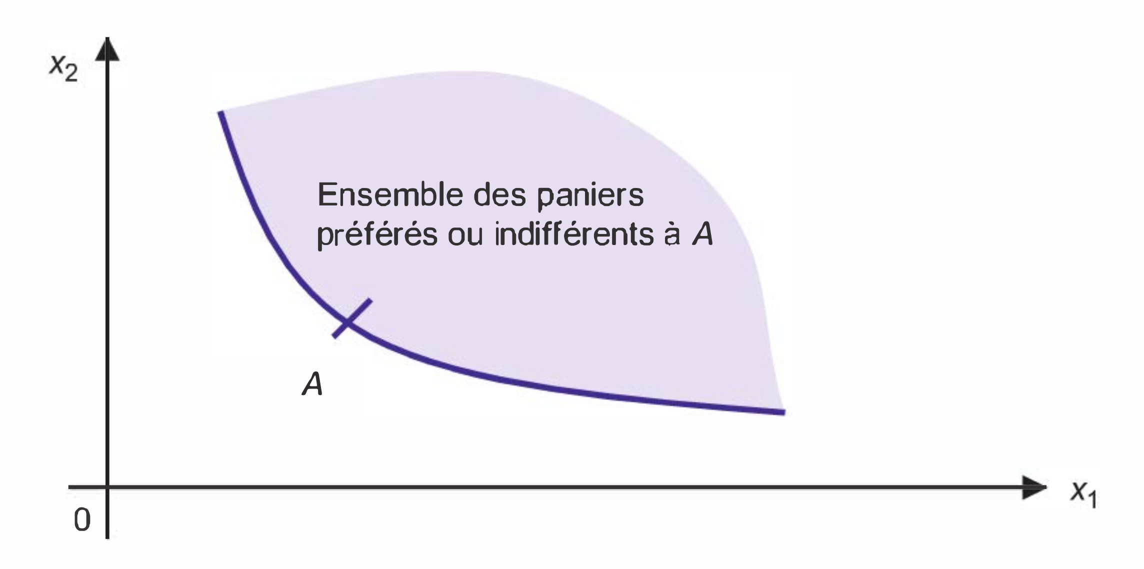


Figure 8

Ainsi, les préférences seront convexes si l'ensemble ci-dessus décrit est lui-même convexe. Dans le graphique ci-dessus, l'ensemble en question est bien convexe car, d'un point de vue mathématique, toute combinaison linéaire convexe de points de cet ensemble appartient elle-même à l'ensemble.

Théorème

Dans le cas de paniers de deux biens, les préférences du consommateur sont (strictement) convexes si et seulement si ses courbes d'indifférence sont (strictement) convexes.

8.2 Interprétation de la convexité des préférences : le goût pour la variété

Comment serait apprécié un panier de biens diversifié en comparaison de 2 paniers de biens qui ne le seraient pas ? Supposons qu'un consommateur soit indifférent entre des paniers *A* et *B* : le panier *A* est un panier qui contient beaucoup de bien 1 et peu de bien 2 ; le panier *B* est un panier qui contient beaucoup de bien 2 et peu de bien 1. Ces deux paniers étant également appréciés par le consommateur, ils apparaissent sur une même courbe d'indifférence. Envisageons maintenant un autre panier, désigné par *C*, qui contient une quantité moyenne de bien 1 et une quantité moyenne de bien 2 (d'un point de vue mathématique, le panier *C* serait construit comme une combinaison linéaire convexe des paniers *A* et *B*), puis traçons la courbe d'indifférence passant par le point *C* (figure 9).

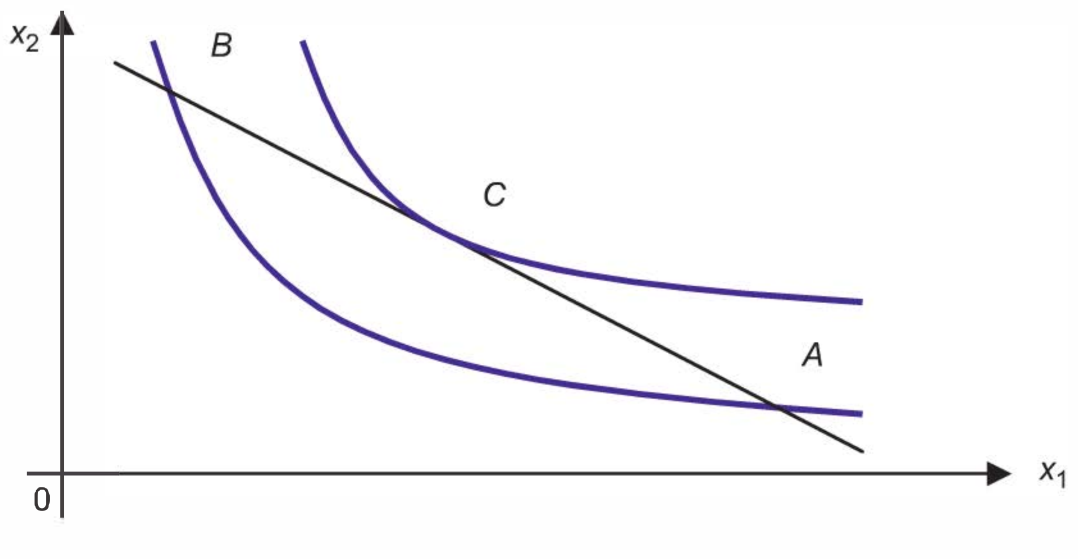


Figure 9

Le panier *C* est situé sur une courbe d'indifférence plus haute que celle passant à la fois par *A* et *B*. Cela signifie que ce panier « moyen » apporte un niveau de satisfaction plus élevé au consommateur. En d'autres termes, le panier diversifié apporte plus de satisfaction que l'un ou l'autre des paniers presque exclusivement composés d'un seul et unique bien : le consommateur préfère la **diversité**, la **variété**.

On interprète la convexité des préférences comme du goût pour la variété, la diversité, les mélanges.

La convexité des préférences explique pourquoi les gens aiment voyager, achètent des produits « exotiques » ou s'intéressent aux films ou à la musique étrangère. Si les préférences des consommateurs, dans leur majorité, n'étaient pas convexes, les gens participeraient peu à des relations marchandes et le commerce international serait embryonnaire. Si nous n'aimions pas la variété, nous n'irions point en vacances pour découvrir d'autres régions ou d'autres pays. Nous nous contenterions de rester dans notre logement et nous mangerions tous les jours les mêmes repas. Nous achèterions toujours le même type de vêtements et de chaussures et nous écouterions toujours la même chanson. Nous lirions toujours le même livre, etc. bref, nous serions extraordinairement renfermés sur nous-mêmes. Il existe néanmoins des cas de préférences non strictement convexes. On peut imaginer que dans un certain nombre de cas, les préférences d'un consommateur sur des paniers de 2 ou plusieurs biens puissent être **concaves**. Ceci révèle de la préférence, non pour la variété, mais pour l'uniformité (comportement de « collectionneur », de préférence géographique, d'addiction, etc). Par exemple, un consommateur préfére n'acheter que des produits alimentaires locaux ou seulement des produits alimentaires exotiques (figure 10).

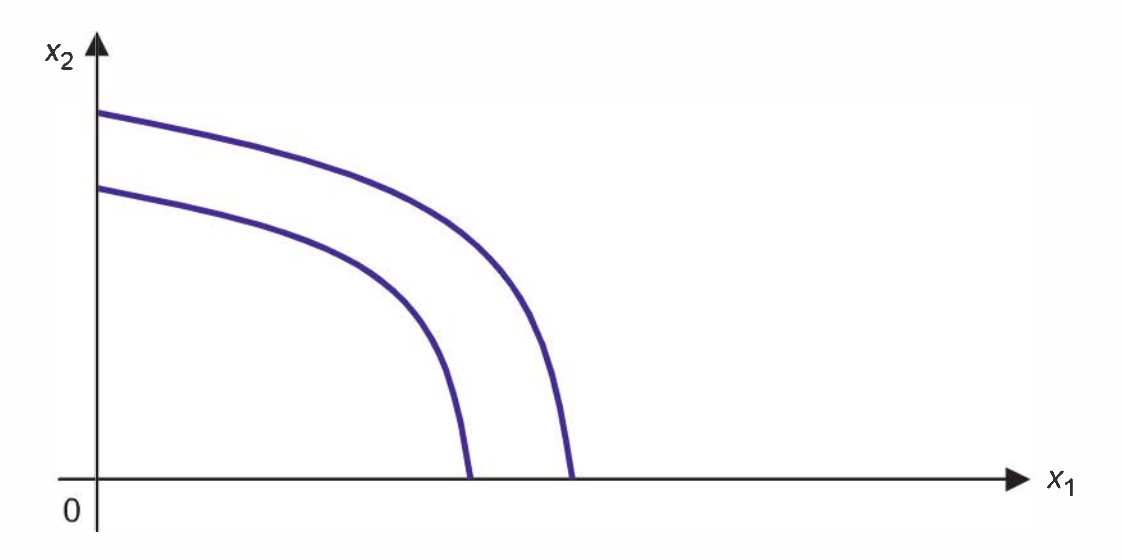


Figure 10 Cas de préférences strictement concaves

8.3 Les biens complémentaires

Des biens complémentaires sont des biens dont la consommation conjointe et en proportions données engendre un niveau de satisfaction déterminé qui ne sera pas accru par la consommation additionnelle isolée d'un quelconque des biens. L'exemple classique est celui des souliers droit et gauche d'une paire de chaussure ou du transistor et de ses piles ou encore du timbre et de l'enveloppe. On peut tracer les courbes d'indifférence relatives à un couple de biens complémentaires comme suit (figure 11).

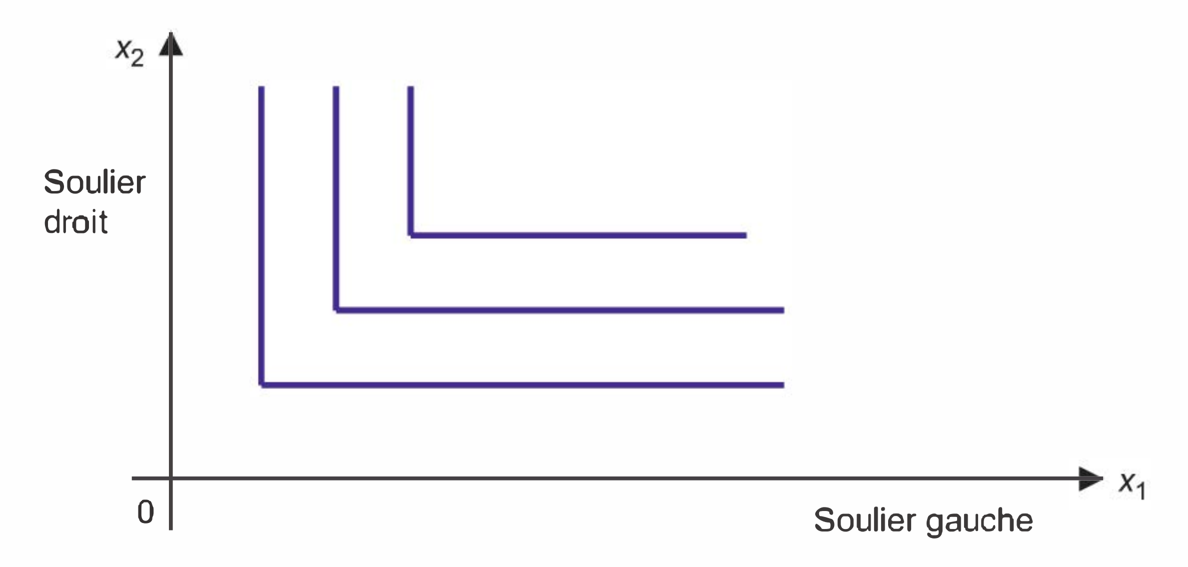


Figure 11 Biens complémentaires

Si l'on dispose d'1 soulier gauche et d'1 soulier droit, on obtient un certain niveau de satisfaction. Si l'on obtient 1 soulier gauche supplémentaire sans obtenir de soulier droit supplémentaire, notre satisfaction sera inchangée (et vice versa si l'on obtient si l'on obtient 1 soulier droit supplémentaire sans obtenir de soulier gauche supplémentaire). En revanche, si l'on obtient simultanément 1 soulier gauche et 1 soulier droit supplémentaires, notre satisfaction va s'accroître brusquement (et nous conduire sur le « coude » supérieur). Remarquons que, dans cet exemple, les courbes d'indifférences devraient être une collection de points (car les biens 1 et 2 sont des biens indivisibles) et non des « courbes » continues (en forme de coudes) comme nous les figurons ici. Les fonctions d'utilité conduisant à cette sorte de courbes d'indifférence sont du type $u(x_1 ; x_2) = Min \{x_1 ; x_2\}$. L'opérateur « Min » permet de traduire l'intuition que nous avons décrite

ci-dessus : la satisfaction éprouvée par le consommateur sera limitée par le nombre d'unité de la sorte de biens la plus faiblement présente dans le panier ; ainsi, un panier composé de 5 souliers gauches et de 3 souliers droits n'apportera pas plus de satisfaction qu'un panier composé de 3 souliers gauches et de 3 souliers. Le cas de biens complémentaires est l'expression d'une préférence ultime pour la variété car, dans de telles circonstances, sans varieté la satisfaction est nulle.

8.4 Les biens parfaitement substituables

Des biens parfaitement substituables sont des biens dont la consommation engendre un niveau de satisfaction déterminé par la quantité globale consommée, indépendamment de la ventilation des quantités consommées entre les différents biens. Un exemple pourrait être celui de trajets en transports urbains dans des bus roulant au bio-carburant (bien 1) ou à l'électricité (bien 2). Traçons ci-dessous les courbes d'indifférence relatives à un couple de biens ou services parfaitement substituables (figure 12).

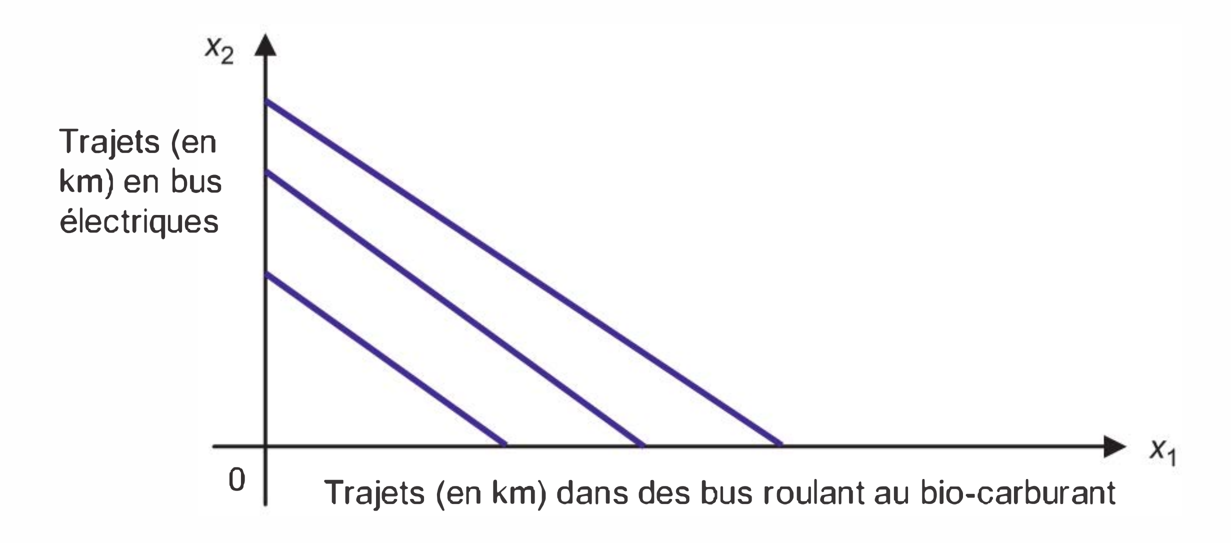


Figure 12 Biens parfaitement substituables

Une telle configuration est presque caricaturale. Considérer comme des biens « différents » des bus utilisant des modes de propulsion différents n'est pas nécessairement pertinent. Cet exemple a pour vertu de faire apparaître le cas polaire du cas de biens complémentaires. Ici, la question de la préférence pour la variété devient « hors de propos » puisque, dans une

acception un peu plus large, les 2 variétés de biens n'en font qu'une seule. Des courbes d'indifférence telles que celles obtenues ci-dessus peuvent l'être à partir de fonctions d'utilité du type $u(x_1 ; x_2) = x_1 + x_2$. L'« additivité » de la fonction d'utilité caractérise les biens parfaitement substituables.

Pour aller plus loin

Démonstration que $\frac{dx_2}{dx_1}$ est égal à $-\frac{Um_1}{Um_2}$

Partons de l'équation de la courbe d'indifférence $u(x_1 ; x_2) = K$ et différencions cette expression :

$$\frac{\partial u(x_1;x_2)}{\partial x_1}dx_1 + \frac{\partial u(x_1;x_2)}{\partial x_2}dx_2 = 0$$

car K est par définition un terme constant. Ceci est équivalent à :

$$\frac{\partial u(x_1;x_2)}{\partial x_1}dx_1 = -\frac{\partial u(x_1;x_2)}{\partial x_2}dx_2$$

$$\Leftrightarrow \frac{dx_2}{dx_1} = -\frac{\frac{\partial u(x_1;x_2)}{\partial x_1}}{\frac{\partial u(x_1;x_2)}{\partial x_2}} = -\frac{Um_1}{Um_2}.$$

2 La décision optimale du consommateur

Mots-clés

Contrainte budgétaire, condition équi-marginale, courbe d'Engel, élasticité, effet substitution et effet revenu.

Les préférences des individus peuvent être caractérisées et modélisées (chapitre 1). Ces préférences intrinsèques, c'est-à-dire exprimées « dans l'absolu », s'expriment le plus souvent dans des circonstances où elles sont confrontées à une contrainte budgétaire. À l'exception de quelques individus très riches qui disposent de tant d'argent qu'une vie entière ne paraît pas suffisante pour le dépenser, chacun effectue, à des degrés divers, des choix budgétairement contraints. C'est la première partie du message de la théorie économique : chacun tente de constituer le meilleur panier de biens ou services possible au regard des ressources limitées dont il dispose. Ce processus d'optimisation, mené par chaque individu, sera longuement détaillé dans ce chapitre et le suivant.

Mais, la théorie économique délivre aussi un second message : chacun, lorsqu'il dépense tout ou partie de ses ressources, se prive d'opportunités extérieures : en achetant un livre de microéconomie, l'étudiant se prive de l'opportunité d'aller plusieurs fois au cinéma ; en achetant un abonnement à une chaîne qui diffuse des matchs de football, le salarié modeste se prive de l'opportunité de remplacer rapidement sa voiture usée ; en achetant une puissante berline allemande, le cadre supérieur se prive de l'opportunité d'acheter prochainement un voilier… Tous nos choix s'apparentent à des sacrifices. Mais, si nous nous comportons rationnellement, nous optons pour le moindre sacrifice : nos choix reflètent la renonciation à la consommation de biens et services la moins psychologiquement « coûteuse »…

1 L'ensemble budgétaire du consommateur

Il faut en priorité formaliser la contrainte budgétaire pesant sur le consommateur. Nous allons donc supposer que le consommateur ne dispose que d'un revenu limité. De leur côté, les biens susceptibles d'être acquis sont caractérisés par un prix unitaire non nul si bien que le consommateur ne peut pas acquérir ces biens en quantité illimitée.

Dans une optique d'équilibre « partiel », nous travaillons en supposant que le prix des biens et le revenu du consommateur sont des données exogènes.

Dans la modélisation d'un phénomène économique, une **variable** est dite **exogène** quand la valeur qu'elle prend est déterminée extérieurement au modèle. Par opposition, une variable dont la valeur se détermine lors de la résolution du modèle est dite **endogène**.

Nous nous positionnons dans une optique d'**équilibre « partiel »** car nous ne nous préoccupons pas de la manière dont se forment les prix des différents biens ou services, ni de la manière dont le revenu du consommateur se détermine. Pour établir simultanément comment se forment, les prix, le revenu et les demandes optimales de bien, il faut se placer dans une perspective d'équilibre « général ». Ce projet ambitieux sera abordé dans le chapitre 5. Dans le présent chapitre, nous nous bornons à supposer que les prix existent et s'imposent au consommateur d'une part, que ce dernier dispose d'un revenu R fixé d'autre part.

Il faut encore préciser que les prix des biens et services sont des prix unitaires définitivement établis. Ainsi, il n'est pas question, dans une première approche, de supposer que le prix serait dégressif en fonction des quantités achetées, ou qu'à l'inverse, à proximité de la rupture de stock, le prix unitaire s'envole : le consommateur peut, à ce prix unitaire donné, acheter le nombre d'unités du bien qu'il souhaite (sous réserve que son revenu le lui permette) : si le consommateur souhaite acquérir 50 casquettes à l'effigie

de son club de football favori (bien que ne disposant que d'un seul « chef » à couvrir), il le peut, sans restriction. Des rationnements sont néanmoins possibles quand le nombre d'unités de biens que chaque consommateur peut acheter est plafonné (voir plus bas).

Les prix et le revenu donnés délimitent l'ensemble des paniers de biens susceptibles d'être acquis par l'individu, ou ce que l'on appelle l'ensemble budgétaire du consommateur. Nous désignons par R le revenu du consommateur et par p_1, p_2, ..., p_n les prix des biens 1, 2, ..., n. L'ensemble des paniers de biens qu'il sera possible au consommateur d'acquérir est délimité par l'inégalité suivante :

$$\text{Dépenses} \leq \text{Ressources}$$

Il est possible de réécrire cette inéquation en exprimant la somme des dépenses en valeur d'une part, les ressources d'autre part.

1.1 Les dépenses

Dans la formalisation développée dans le précédent chapitre, les variables (endogènes) x_1, x_2, ..., x_n désignent le nombre d'unités de chacun des biens que le consommateur acquiert. Puisque p_1 désigne le prix unitaire du bien 1, la dépense consacrée à l'acquisition de la quantité x_1 de ce bien sera $p_1 \times x_1$. De même puisque p_2 désigne le prix unitaire du bien 2, la dépense consacrée à l'acquisition de la quantité x_2 de ce bien sera $p_2 \times x_2$, ...et ainsi de suite jusqu'au bien n. La dépense totale consacrée à l'acquisition des n biens (ou dépense en valeur) sera donc :

$$p_1 \times x_1 + p_2 \times x_2 + ... + p_n \times x_n$$

Remarque : Il est convenu de ne pas noter le signe de la multiplication apparaissant entre des variables au sein d'une équation (ou d'une inéquation). Nous opterons, dans toute la suite, pour cette convention. Ainsi, nous noterons la dépense totale en valeur $p_1x_1 + p_2x_2 + ... + p_nx_n$.

1.2 Les ressources

Comme indiqué plus haut, la seule ressource dont dispose le consommateur est son revenu R.

1.3 La contrainte budgétaire

En réunissant les deux termes de l'inéquation caractérisant l'ensemble des paniers de biens susceptibles d'être acquis par le consommateur, nous obtenons la contrainte budgétaire du consommateur :

$$p_1x_1 + p_2x_2 + \ldots + p_nx_n \leq R$$

La **contrainte budgétaire** du consommateur est l'inégalité qui caractérise l'ensemble des paniers de biens que le consommateur est susceptible d'acquérir étant donné son revenu et les prix auxquels les biens sont vendus. Dans le cas d'un panier de n biens, elle s'écrit :

$$p_1 x_1 + p_2 x_2 + \ldots + p_n x_n \leq \mathrm{R}$$

Comme précédemment, c'est dans le cas où $n = 2$ biens que nous allons pouvoir facilement proposer une représentation graphique de cette contrainte budgétaire. En effet, dans ce cas, la contrainte budgétaire se limite à :

$$p_1 x_1 + p_2 x_2 \leq R$$

Il est possible de réécrire cette inéquation comme : $x_2 \leq \frac{R}{p_2} - \frac{p_1}{p_2} x_1$.

Ceci délimite une zone triangulaire dans le plan $(0\,;x_1\,;x_2)$. Cette zone est qualifiée de zone budgétairement accessible ou **ensemble budgétaire du consommateur**. Elle est délimitée inférieurement par les axes et supérieurement par une droite, la droite d'équation $x_2 = \frac{R}{p_2} - \frac{p_1}{p_2} x_1$, encore appelée droite de budget. La droite de budget est donc une droite de pente $-\frac{p_1}{p_2}$ (c'est-à-dire une droite décroissante), dont l'ordonnée à l'origine est $\frac{R}{p_2}$ et l'abscisse à l'origine est $\frac{R}{p_1}$ (figure 1).

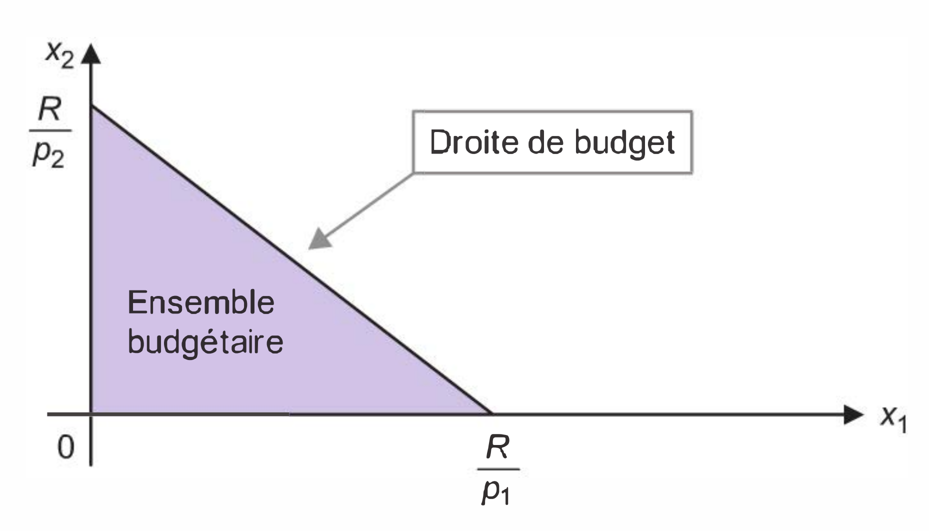


Figure 1 Ensemble budgétaire du consommateur

Les axes sont-ils inclus dans l'ensemble budgétaire ? À quoi correspondrait le choix d'un panier de deux biens positionné sur l'un des axes ? Lorsqu'un panier est positionné sur l'axe des abscisses (0 ; x_1), cela signifie que la quantité consommée du bien 2 est nulle ; réciproquement, lorsqu'un panier est positionné sur l'axe des ordonnées (0 ; x_2), cela signifie que la quantité consommée du bien 1 est nulle. Est-il raisonnable d'envisager que la quantité consommée de l'un des biens du panier soit nulle ? C'est non seulement raisonnable, mais même souhaitable pour être en mesure de traiter d'applications pertinentes. Si l'on raisonne dans une optique assez large, par exemple une optique dans laquelle les paniers comportent des biens que l'on ne consomme qu'épisodiquement au cours de sa vie (par exemple des biens durables : biens immobiliers, automobiles, meubles, etc.), il est indispensable que la quantité consommée d'un bien lors d'une période particulière (par exemple, une année) puisse être nulle. Mais ce n'est pas toujours l'hypothèse retenue dans le cas de paniers de biens de consommation courante, lorsque l'on raisonne sur un horizon de temps court. L'hypothèse généralement retenue est alors celle de **biens nécessaires**.

Un **bien** est **nécessaire** si, dès que le revenu est strictement positif, la quantité consommée de ce bien est strictement positive (même si elle est infime).

Pour formaliser l'hypothèse qu'un bien *i* est nécessaire, on écrira $x_i > 0$.

L'intérêt de cette hypothèse est principalement technique. Elle conduit à ce que la décision optimale du consommateur, si elle existe, soit toujours une solution dite *intérieure* (par opposition à une solution *en coin*), c'est-à-dire une solution qui ne se situe jamais sur un des axes. Nous reviendrons plus loin sur cette question.

■ Contraction ou extension de l'ensemble budgétaire

Si le revenu du consommateur ou le prix de l'un des biens se modifie, l'ensemble budgétaire est modifié lui aussi : il se contracte quand le prix d'un bien augmente ou le revenu diminue, il s'étend quand le prix d'un bien diminue ou le revenu augmente.

□ Effet d'une variation de revenu

Supposons que le revenu augmente : il passe de R à R' (où $R' > R$). La pente de la droite de budget est inchangée ; en revanche ses abscisse et ordonnée à l'origine augmentent. La figure 2 permet de visualiser la manière dont l'ensemble budgétaire s'étend.

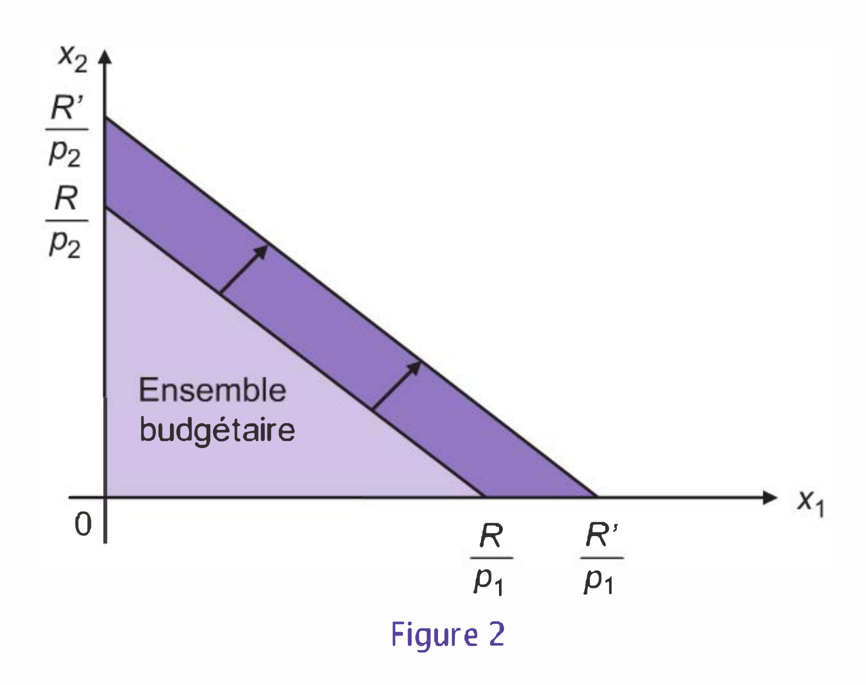


Figure 2

□ Effet d'une variation de prix

Supposons maintenant que le prix du bien 1 augmente : le prix passe de p_1 à p_1' (où $p_1' > p_1$). La droite de budget devient plus pentue (sa pente augmente en valeur absolue). Son abscisse à l'origine diminue. Son ordonnée à

l'origine est inchangée puisque ni le revenu, ni le prix du bien 2 ne sont supposés se modifier. La figure 3 permet de visualiser la manière dont l'ensemble budgétaire se contracte.

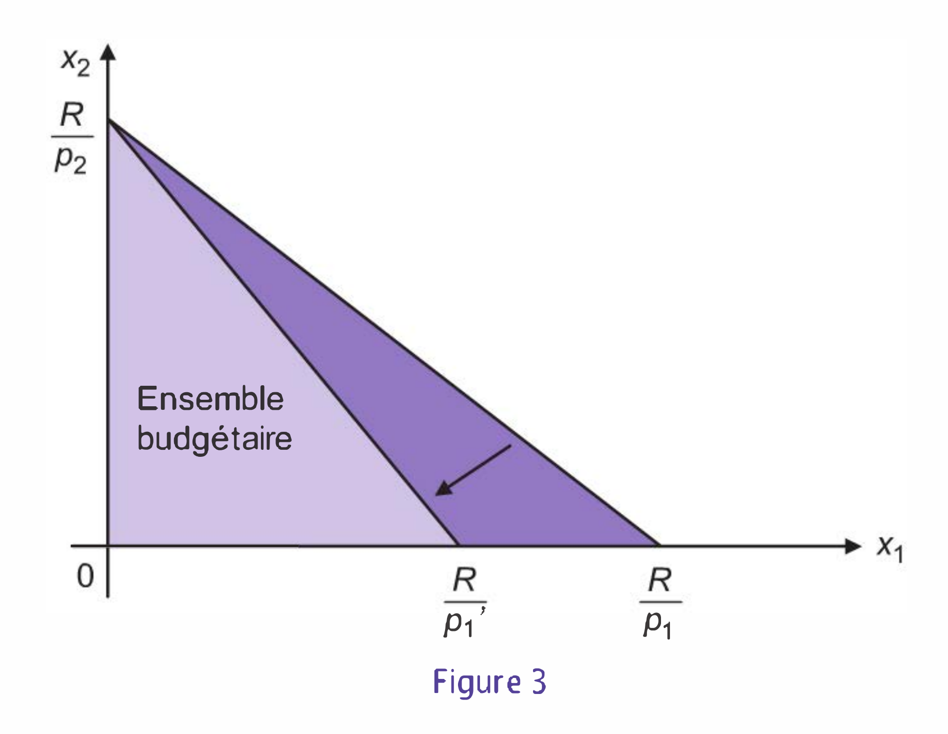


Figure 3

☐ Effet de l'instauration d'un rationnement

Dans certaines circonstances, le consommateur ne peut pas acheter autant d'unités de bien qu'il le souhaite. Par exemple, en situation de guerre ou de pénurie, certains biens alimentaires (pain, viande, sucre, etc.) peuvent être temporairement rationnés. De manière moins dramatique, dans certaines grandes surfaces, en cas de promotions très avantageuses sur des bouteilles de champagne ou de pastis, il arrive que le nombre de bouteilles que chaque famille puisse acheter soit limité (par exemple 5 ou 10). Dans de tels cas, l'ensemble budgétaire du consommateur est réduit. Par exemple, si la quantité maximale de bien 2 que chaque individu est autorisé à acquérir est limitée à k_2, on écrit dans le programme du consommateur $x_2 \leq k_2$, et la partie de l'ensemble budgétaire située au dessus de la droite horizontale d'équation $x_2 = k_2$, devient inaccessible. La figure 4 permet de visualiser la manière dont l'ensemble budgétaire se contracte.

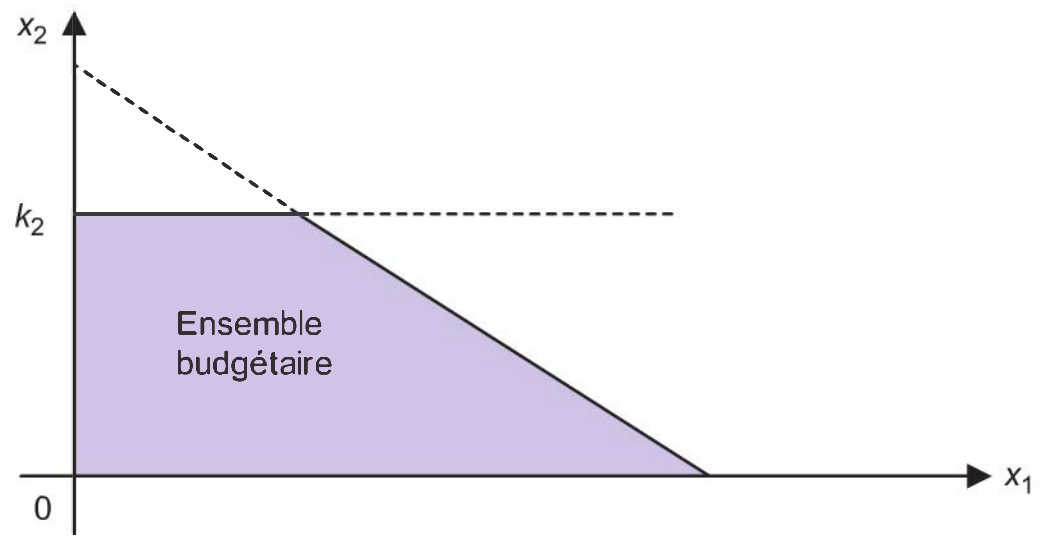


Figure 4

2 Résolution du problème de décision optimale du consommateur

Nous connaissons maintenant les deux ingrédients à partir desquels va se construire le choix du consommateur : d'une part ses préférences (qu'incarnent les courbes d'indifférence), d'autre part sa contrainte budgétaire (matérialisée par l'ensemble budgétaire). En confrontant ces deux ingrédients, nous allons illustrer graphiquement l'arbitrage auquel il est confronté : comment peut-il répartir son revenu (limité) entre les différents biens du panier, de manière à ce que le niveau de satisfaction atteint soit le plus élevé possible ?

Sur la figure 5, nous faisons apparaître la droite de budget et le faisceau des courbes d'indifférence du consommateur : à chaque courbe d'indifférence correspond un niveau particulier de satisfaction et ce niveau de satisfaction est d'autant plus grand que la courbe est située vers le haut et la droite.

Où se situe la décision optimale du consommateur ? Tant qu'il existe un « point de contact » d'une courbe d'indifférence avec l'ensemble budgétaire, il est possible d'atteindre cette courbe et donc ce niveau de satisfaction. En rappelant que la droite de budget elle-même appartient à

l'ensemble budgétaire du consommateur, il apparaît que la plus haute courbe d'indifférence compatible avec l'ensemble budgétaire est celle qui apparaît sur la figure 6.

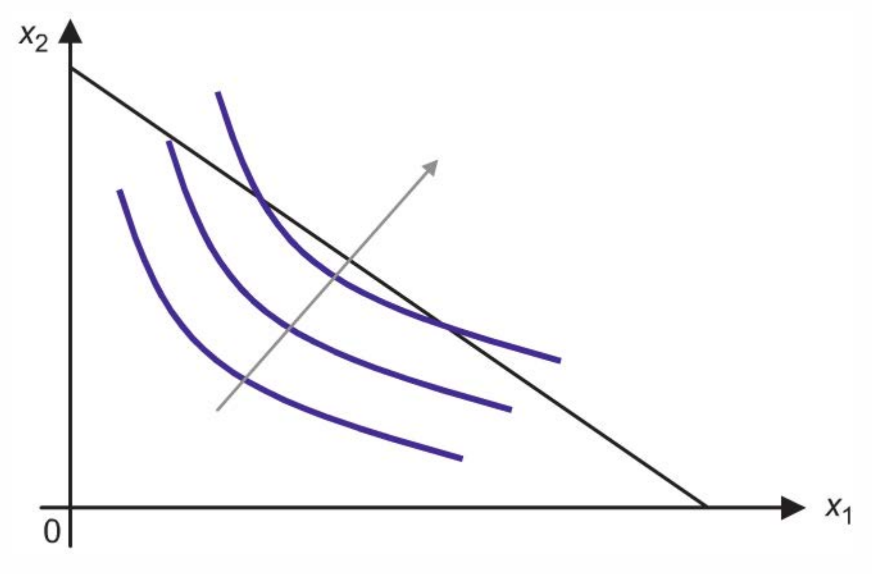


Figure 5

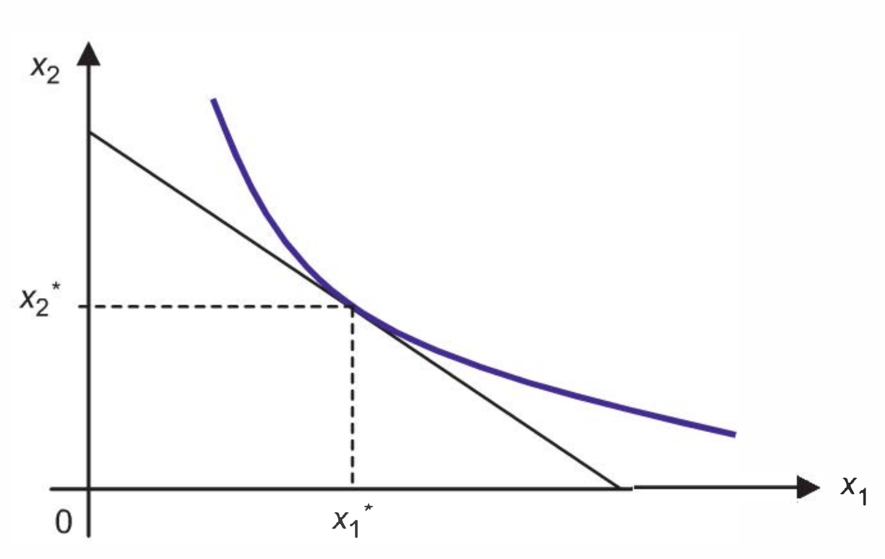


Figure 6 Décision optimale du consommateur

Le panier de biens (x_1* ; x_2*) est l'optimum, c'est-à-dire le panier de biens qui apporte le plus haut niveau de satisfaction possible au consommateur (étant donné les prix des biens et son revenu). Nous remarquons qu'à l'optimum, la courbe d'indifférence est précisément *tangente* à la droite de budget, ou, en d'autres termes, que la tangente à la courbe d'indifférence en ce point est confondue avec la droite de budget. Les pentes des deux droites (la tangente et la droite de budget) sont donc égales. Rappelons que

la pente de la tangente à la courbe d'indifférence est égale au *taux marginal de substitution* entre les biens et que la pente de la droite de budget est égale à $-\frac{p_1}{p_2}$. Ainsi, à l'optimum :

$$TmS_{1/2} = -\frac{p_1}{p_2}$$

Caractérisation de la décision optimale du consommateur (1)

La décision optimale du consommateur est un panier de biens $(x_1^* ; x_2^*)$ tel que le taux marginal de substitution entre les biens soit égal au rapport de leurs prix, encore appelé *prix relatif* (du bien 1 relativement au bien 2).

La première manière d'interpréter cette propriété est d'indiquer que la décision optimale sera atteinte quand le consommateur sera parvenu à faire coïncider le ratio qui mesure les proportions dans lesquelles il est prêt à substituer un bien par un autre avec le ratio des prix (qui exprime, en quelque sorte, les proportions dans lesquelles les conditions de commercialisation des biens « autorisent » la substitution d'un bien par un autre). Il s'agit d'égaliser un ratio de substitution « subjectif » (le TmS) et un ratio de substitution « objectif » (le rapport des prix).

En rappelant que $TmS_{1/2} = -\frac{Um_1}{Um_2}$, nous pouvons aussi écrire que, à l'optimum :

$$\frac{Um_1}{Um_2} = \frac{p_1}{p_2}$$

ou, ce qui est équivalent :

$$\frac{Um_1}{p_1} = \frac{Um_2}{p_2}$$

Cette seconde écriture est appelée « condition équi-marginale ». Elle repose sur les termes $\frac{Um_i}{p_i}$ $(i = 1, 2, \ldots, n)$ que l'on peut interpréter comme mesurant le « surcroît de satisfaction (par euro dépensé) obtenu grâce à la dernière unité consommée du bien i »

Caractérisation de la décision optimale du consommateur (2) : condition équi-marginale

La décision optimale du consommateur est un panier de biens $(x_1^* \,;\, x_2^*)$ tel que le surcroît de satisfaction (par euro dépensé) obtenu grâce à la dernière unité consommée du bien 1 soit égal au surcroît de satisfaction par euro dépensé obtenu grâce à la dernière unité consommée du bien 2.

Si la condition équi-marginale n'était pas vérifiée, le consommateur réorganiserait ses dépenses et parviendrait à augmenter sa satisfaction. Par exemple, si le la dernière unité consommée du bien 1 apportait plus de satisfaction (par euro dépensé) que la dernière unité consommée du bien 2, le consommateur achèterait un peu plus de bien 1 et un peu moins de bien 2 (en rappelant qu'il ne peut pas acquérir simultanément plus des deux biens car il est contraint par son revenu limité). Cet ajustement continuerait tant que les satisfactions marginales (par euro dépensé) demeureraient différentes.

L'hypothèse de convexité stricte des préférences, présentée au chapitre précédent, est une hypothèse décisive pour parvenir au résultat ci-dessus : le panier optimal est un panier qui contient des quantités strictement positives du bien 1 et du bien 2, c'est-à-dire un panier qui est composé de chacun des biens (ce que l'on appelle une **solution intérieure**). En l'absence de l'hypothèse de convexité stricte des préférences, la condition équi-marginale n'est pas nécessairement vérifiée à l'optimum. C'est ce que nous détaillerons maintenant, dans quatre configurations pour lesquelles l'hypothèse de convexité stricte des préférences n'est pas vérifiée. Enfin, dans une cinquième configuration, nous constaterons que la condition équi-marginale peut aussi ne pas être vérifiée en cas de présence d'un rationnement, même lorsque les préférences sont strictement convexes.

2.1 Cas de préférences ni convexes, ni concaves

On peut imaginer que les courbes d'indifférence incarnant les préférences d'un individu sur des paniers de deux biens particuliers prennent la forme suivante (figure 7).

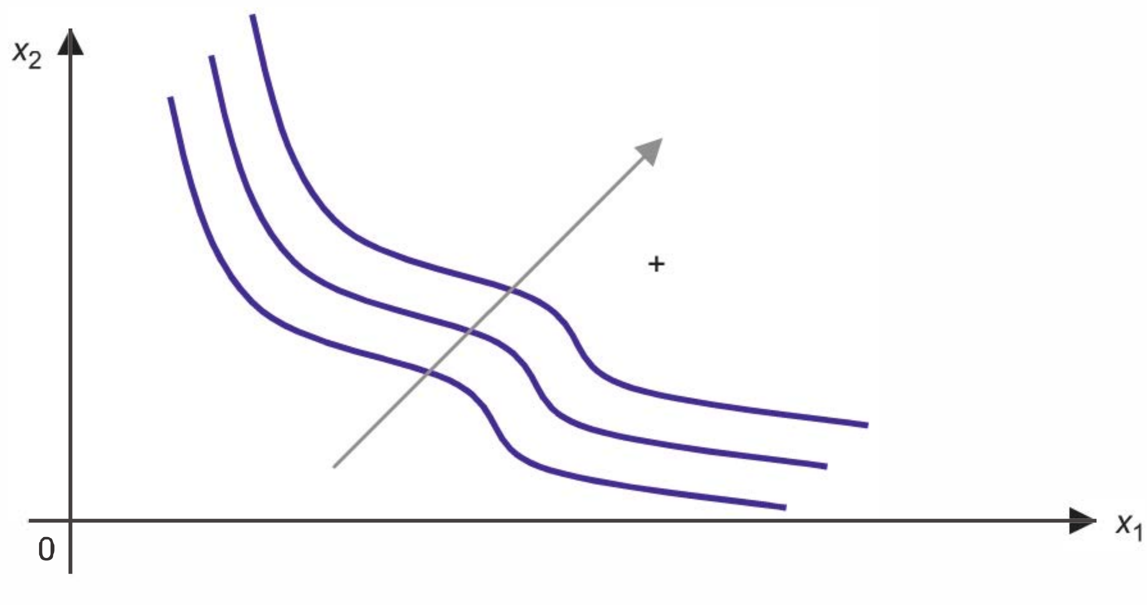


Figure 7

On remarque que les courbes d'indifférence sont bien décroissantes, ne se croisent pas et correspondent à des niveaux d'utilité d'autant plus élevés que l'on se situe plus haut, vers la droite. En revanche, elles ont toutes une partie convexe, puis une partie concave, puis à nouveau une partie convexe. On peut imaginer des biens pour lesquels un comportement de « collectionneur » se déclarerait entre certains seuils de quantités consommées. Par exemple, dès lors dès que le consommateur a acheté 2 ou 3 opéras de Mozart (bien 1), il souhaite par-dessus tout compléter sa collection en acquérant d'autres œuvres de Mozart. Réciproquement, dès lors qu'il a acheté 1 ou 2 opéras d'Offenbach (bien 2), il souhaite par-dessus tout élargir sa culture de l'opéra-bouffe. Confrontons son faisceau de courbes d'indifférence à sa contrainte budgétaire et faisons apparaître la plus haute courbe compatible avec son ensemble budgétaire (figure 8).

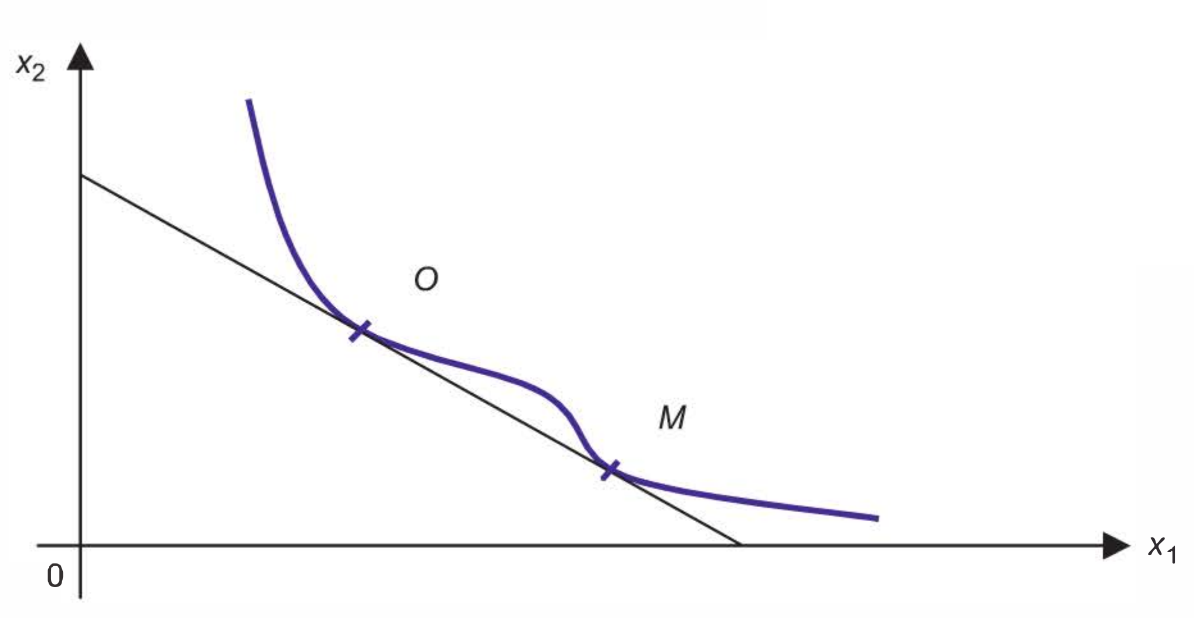


Figure 8

On voit apparaître sur la figure 8 la possible existence de 2 optima ayant en commun de vérifier la condition équi-marginale (dans les 2 cas, la tangente à la courbe d'indifférence est confondue avec la droite de budget). Les 2 optima sont, en termes de satisfaction, équivalents aux yeux du consommateur. Mais, cette configuration pour laquelle émergent 2 optima (entre lesquels il est impossible de discriminer) demeure peu probable. En règle générale, il n'existera qu'un seul optimum (pour lequel la condition équi-marginale sera vérifiée) (figure 9) :

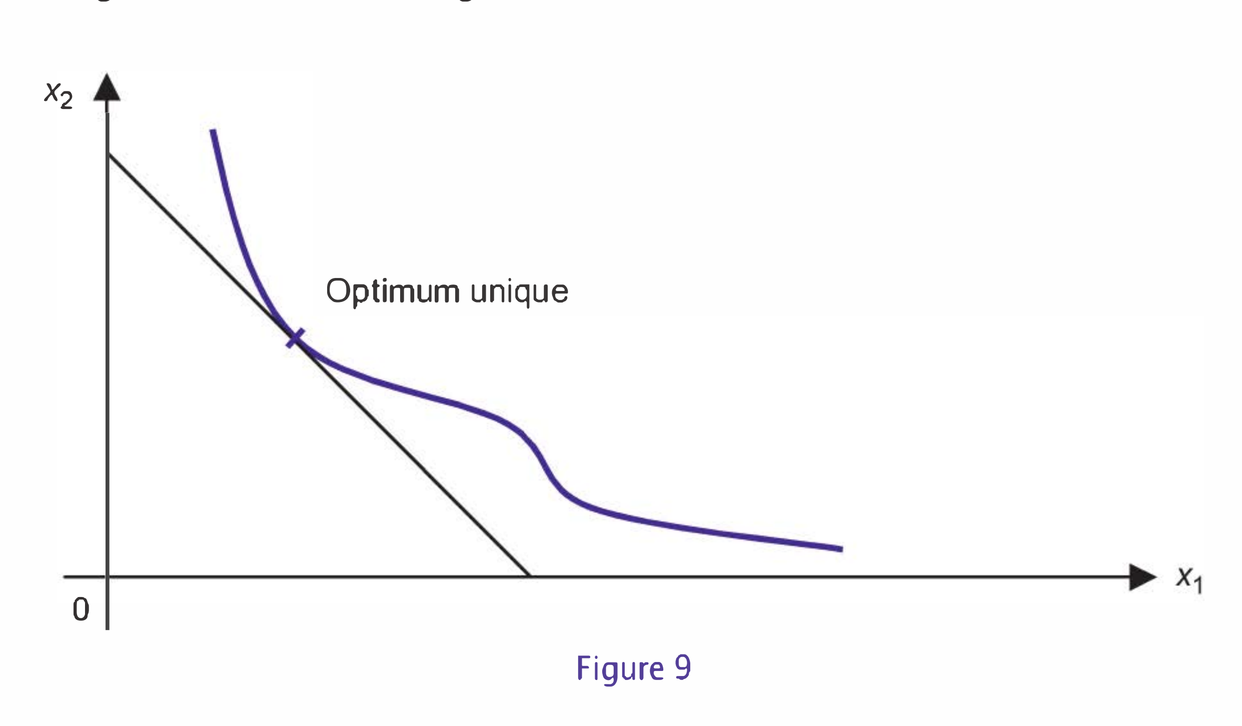


Figure 9

2.2 Cas de préférences strictement concaves

Dans le cas de préférences strictement concaves (comportements d'addiction), il se peut que l'optimum du consommateur soit une **solution en coin**. C'est ce qui apparaît sur la figure 10.

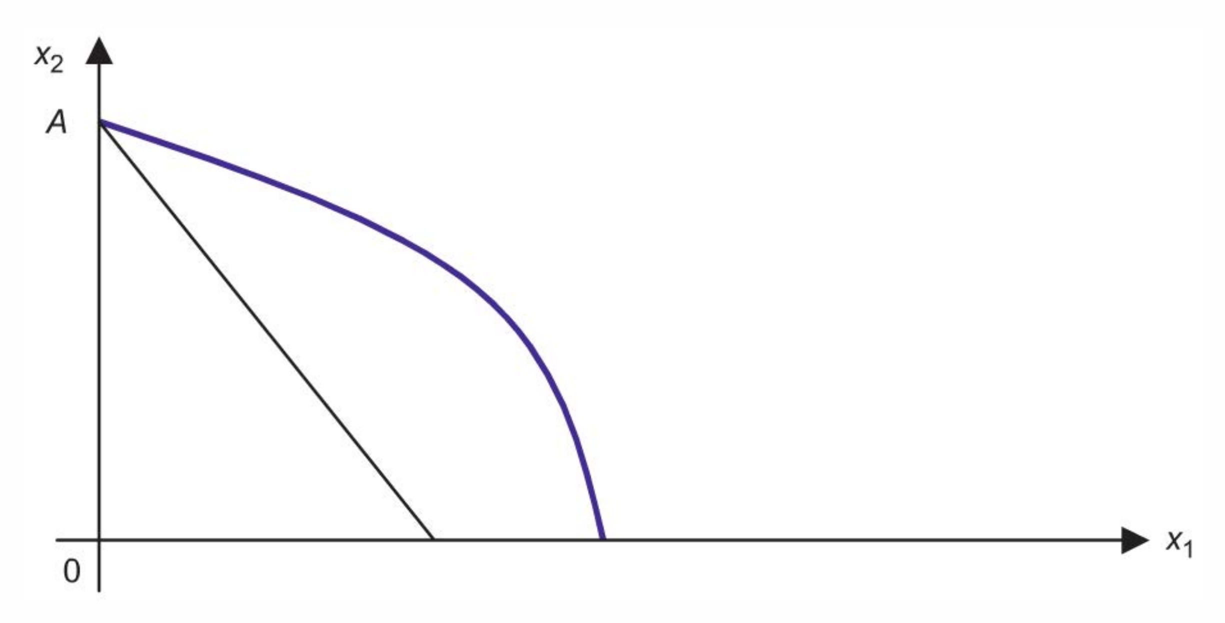


Figure 10

Pour que le point *A* soit en effet l'optimum du consommateur, il faut que le bien 1 soit, dans le cas présent, « non nécessaire ». Dans de telles circonstances, le consommateur consacre tout son revenu à la consommation de bien 2 (on obtient que $(x_1^*; x_2^*) = (0 ; R/p_2)$). Ce serait, par exemple, le choix d'un consommateur de drogue (bien 2), dont la dépendance aux stupéfiantes substances le conduirait à négliger toute autre acquisition. Nous remarquons que la condition équi-marginale n'est pas vérifiée pour l'optimum *A* : la pente de la tangente à la courbe d'indifférence au point *A* n'est plus identique à celle de la droite de budget. Le consommateur ne cherche plus à équilibrer ses acquisitions pour faire coïncider les « surcroîts de satisfaction par euro dépensé », mais se concentre sur l'acquisition exclusive d'un seul et unique bien (en raison de son addiction maladive). En examinant une telle configuration, on est tenté d'objecter qu'il faut néanmoins que ce consommateur acquiert quelque nourriture pour survivre. Ceci correspond, en quelque sorte, à l'idée que la nourriture (qui serait alors le bien 1 dans l'exemple ci-dessus) est un bien « nécessaire ». Si nous

choisissions cette hypothèse, nous serions, en vérité, fort embarrassés : nous tomberions dans une impasse mathématique se traduisant par l'inexistence de solution au problème du consommateur !

2.3 Cas de biens complémentaires

Le cas des biens complémentaires est sans doute le cas le plus facile à traiter graphiquement. En effet, comme nous le constatons sur la figure 11, ce cas offre la garantie de l'existence et de l'unicité de l'optimum.

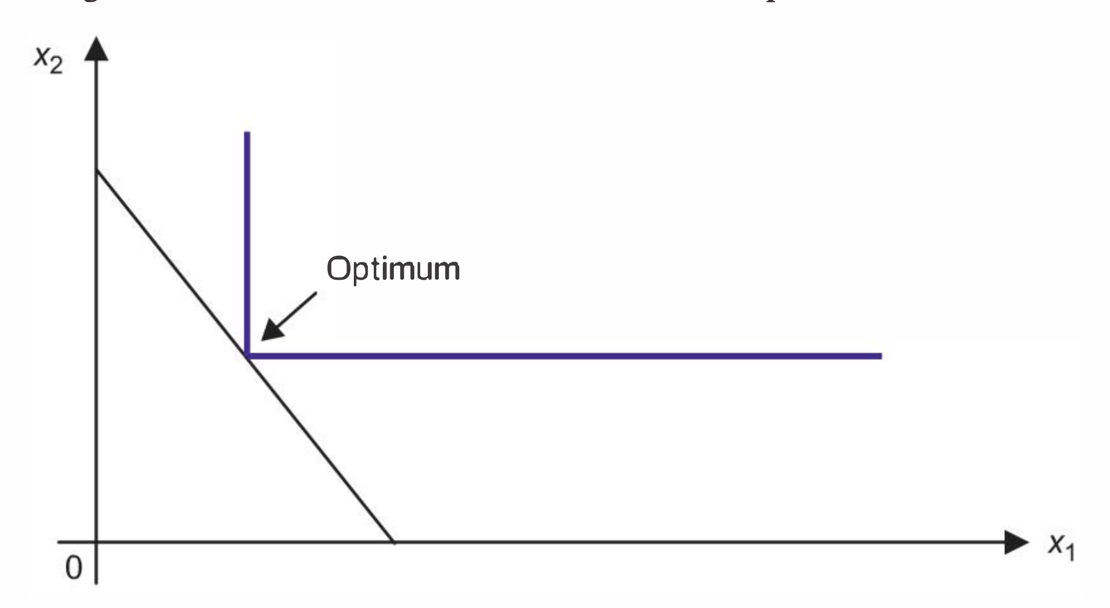


Figure 11

En revanche, la condition équi-marginale n'est pas satisfaite : on ne peut, en effet, pas définir la tangente à la courbe d'indifférence en un de ces points que l'on appelle « coude » (car mathématiquement, en de tels points, la dérivée à droite et la dérivée à gauche ne sont pas égales). Il est donc vain de vouloir vérifier la condition d'optimalité générique.

2.4 Cas de biens parfaitement substituables

Dans le cas de biens parfaitement substituables, nous retrouvons les questionnements liés à l'hypothèse de biens « nécessaires ». Il faut lever cette hypothèse si nous voulons obtenir un optimum, qui sera, sauf cas fortuit, une solution en coin. La figure 12 illustre cette issue.

Pour que le point *B* soit en effet l'optimum du consommateur, il faut que le bien 2 soit, dans cette situation, « non nécessaire ». Si tel est le cas, le

consommateur consacre tout son revenu à la consommation de bien 1 (on obtient alors que $(x_1^*; x_2^*) = (R/p_1 ; 0)$) et la condition équi-marginale n'est pas vérifiée à l'optimum.

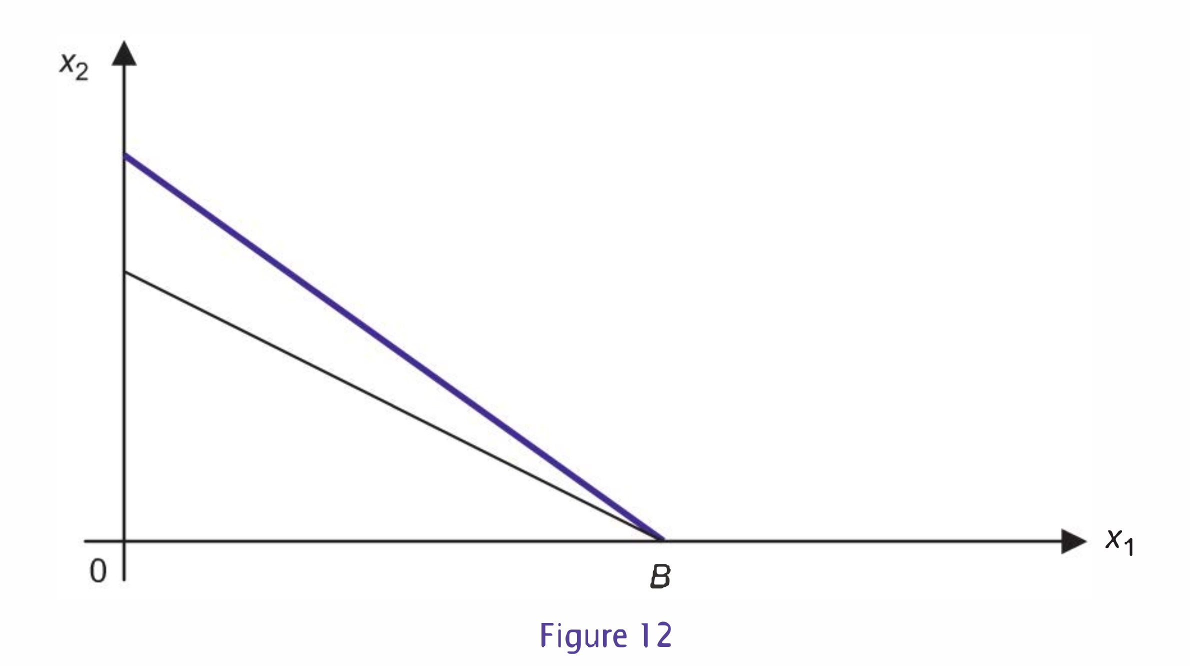


Figure 12

Il se peut fortuitement que les conditions auxquelles sont commercialisés les biens soient telles que le rapport de leurs prix $-p_1/p_2$ soit exactement égal (au signe prêt) au rapport des utilités marginales. Dans un tel cas, la plus haute courbe d'indifférence compatible avec l'ensemble budgétaire est précisément confondue avec la droite de budget. Il y a alors une infinité de solutions (tous les paniers appartenant à la droite de budget) et la condition équi-marginale est vérifiée.

Enfin, en règle générale, dans le cas de biens « nécessaires », il n'existe aucune solution au problème du consommateur.

2.5 Cas de la présence d'un rationnement

Dans le cas d'un rationnement, l'ensemble budgétaire est réduit et prend la forme d'un trapèze. Deux types d'optima du consommateur peuvent exister, en fonction de l'intensité du rationnement :

- soit le rationnement est peu prononcé, auquel cas le consommateur ne bute pas sur l'impossibilité de consommer une quantité de bien supérieure au seuil autorisé ; dans ce cas, la condition équi-marginale

est vérifiée et l'on retrouve les caractéristiques habituelles de la décision optimale du consommateur (figure 13).

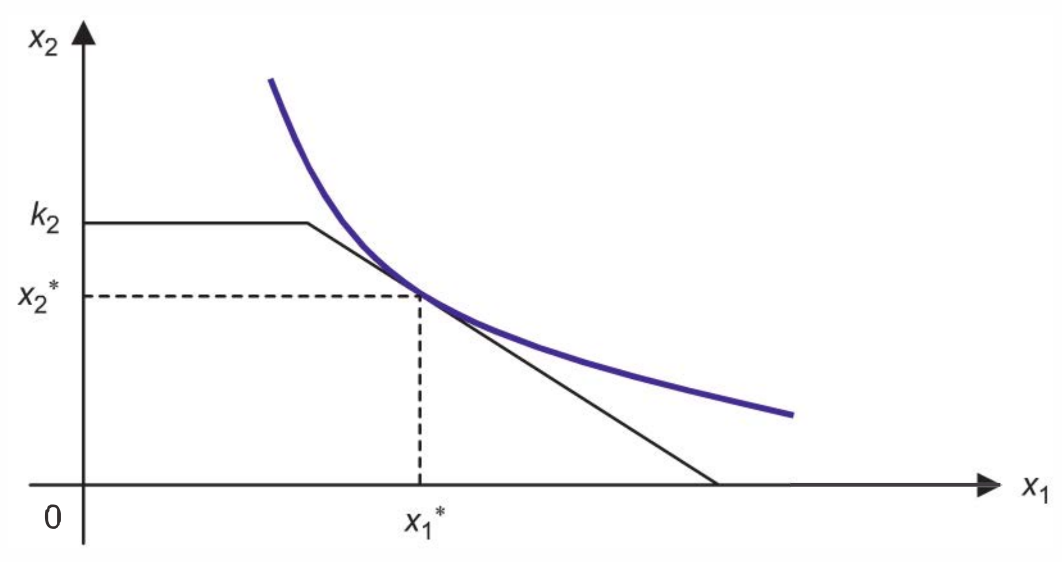


Figure 13

- soit le rationnement est fort, auquel cas le consommateur bute sur cette contrainte quantitative ; dans ce cas, la condition équi-marginale n'est plus vérifiée. Le consommateur choisit de consommer la quantité maximale autorisée du bien soumis à rationnement et alloue le solde de ses ressources à la consommation de l'autre (ou des autres) bien(s) (figure 14).

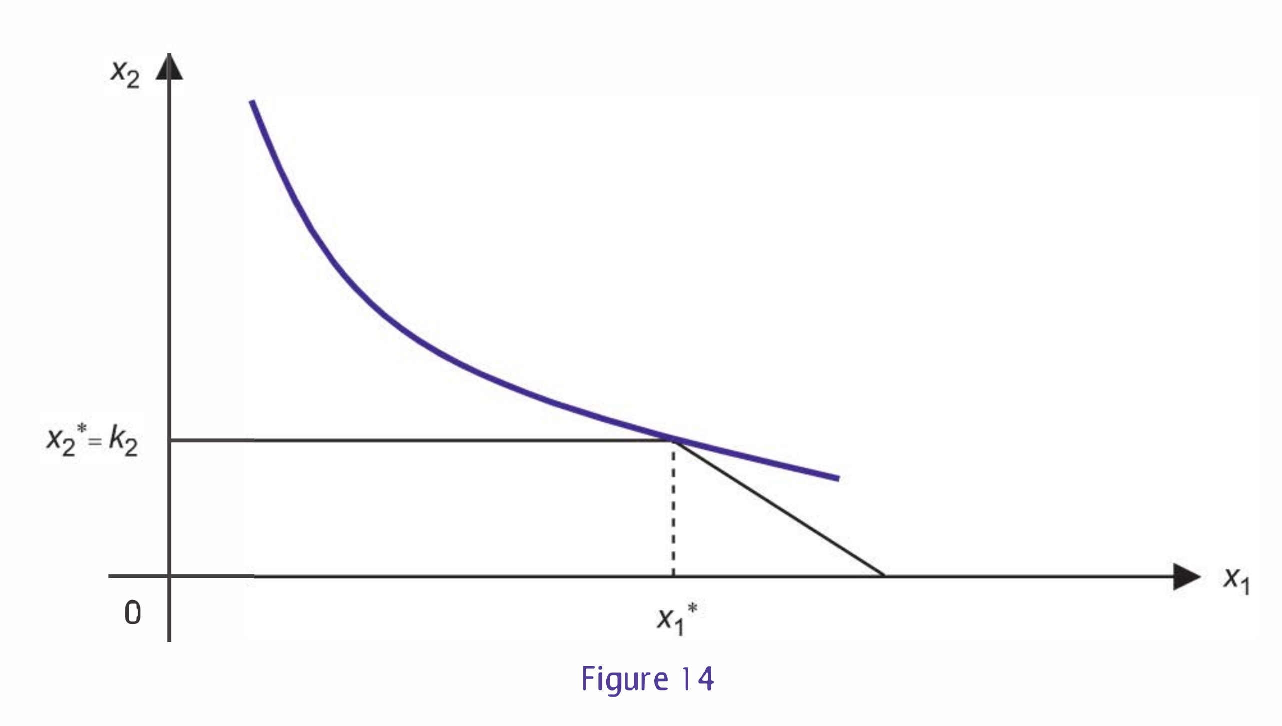


Figure 14

Pour aller plus loin

Résolution du programme du consommateur

Le problème que nous avons résolu graphiquement ci-dessus peut être posé de manière formelle. La recherche du plus haut niveau de satisfaction possible correspond à la maximisation de la fonction d'utilité du consommateur. Cette maximisation se fait sous la contrainte budgétaire présentée plus haut. Ainsi, le consommateur doit résoudre le programme suivant :

$$\begin{array}{ll} \underset{x_1, x_2, \ldots, x_n}{Max} & u(x_1; x_2; \ldots; x_n) \\ \text{sous contrainte} & p_1 x_1 + p_2 x_2 + \ldots + p_n x_n \leq R \end{array}$$

Dans le cas de paniers de deux biens, en considérant que le consommateur alloue tout son revenu disponible à la consommation des biens (le consommateur respecte l'axiome de non-saturation : tant qu'il lui reste le moindre centime d'euro, il souhaite continuer à acquérir des biens), le programme s'écrit :

$$\begin{array}{ll} \underset{x_1, x_2}{Max} & u(x_1; x_2) \\ \text{sous contrainte} & p_1 x_1 + p_2 x_2 = R \Leftrightarrow x_2 = \frac{R}{p_2} - \frac{p_1}{p_2} x_1 \end{array}$$

qui peut être résolu en : $\underset{x_1}{Max} \quad u(x_1; \frac{R}{p_2} - \frac{p_1}{p_2} x_1)$

On obtient alors le couple de consommation optimale $(x_1^*; x_2^*)$ qui vérifie :

$$TMS_{1/2} = -\frac{\frac{\partial u}{\partial x_1}}{\frac{\partial u}{\partial x_2}} = -\frac{p_1}{p_2}$$

3 Impact d'une variation de revenu

Comment le consommateur modifie-t-il son panier de consommation optimal lorsque son revenu disponible augmente ? Un accroissement du revenu conduisait à une extension de l'ensemble budgétaire. Confrontons cette extension de l'ensemble budgétaire au faisceau des courbes d'indifférence. Nous identifions alors les paniers de consommation optimaux relatifs à différents niveaux de revenu (figure 15).

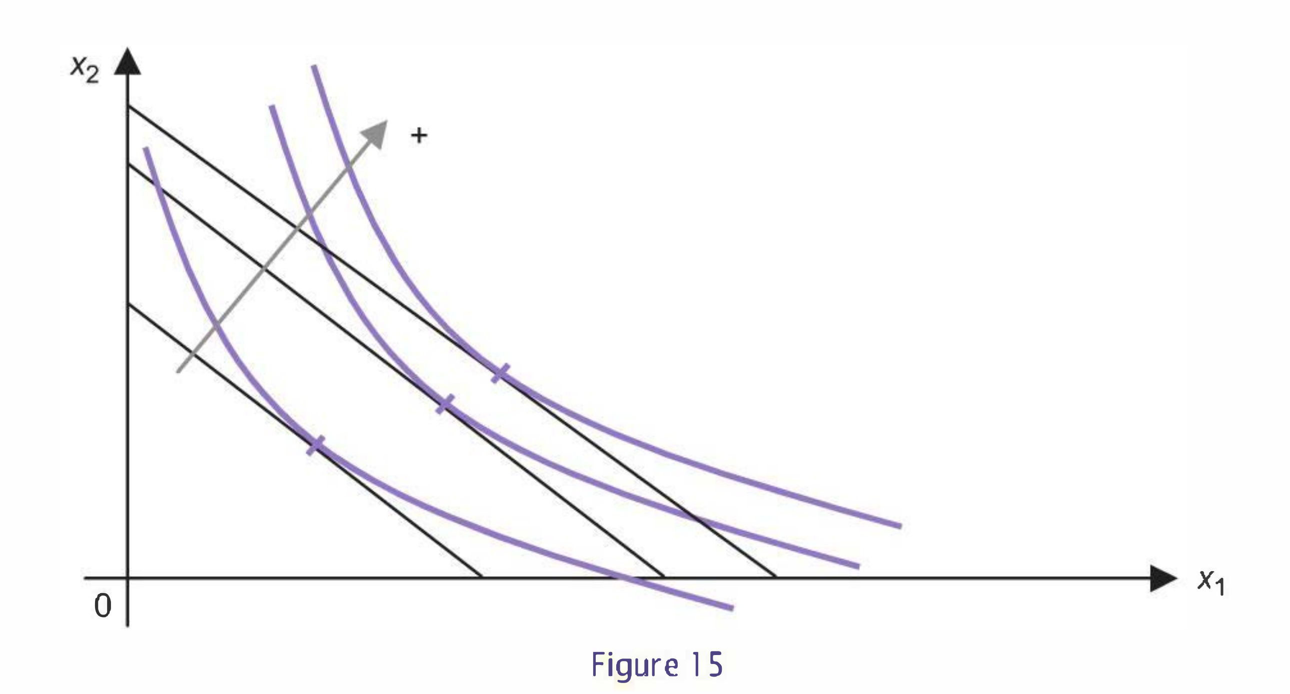


Figure 15

3.1 Chemin d'expansion du revenu

Si nous relions ces différents points, nous obtenons une courbe qui décrit la trajectoire de consommation optimale lorsque le revenu progresse. Cette courbe est appelée **chemin d'expansion du revenu** ou courbe consommation-revenu (figure 16).

La courbe qui, dans le repère $(0 ; x_1 ; x_2)$, décrit les paniers de biens optimaux $(x_1^* ; x_2^*)$ pour les différents niveaux du revenu R est appelée **chemin d'expansion du revenu.**

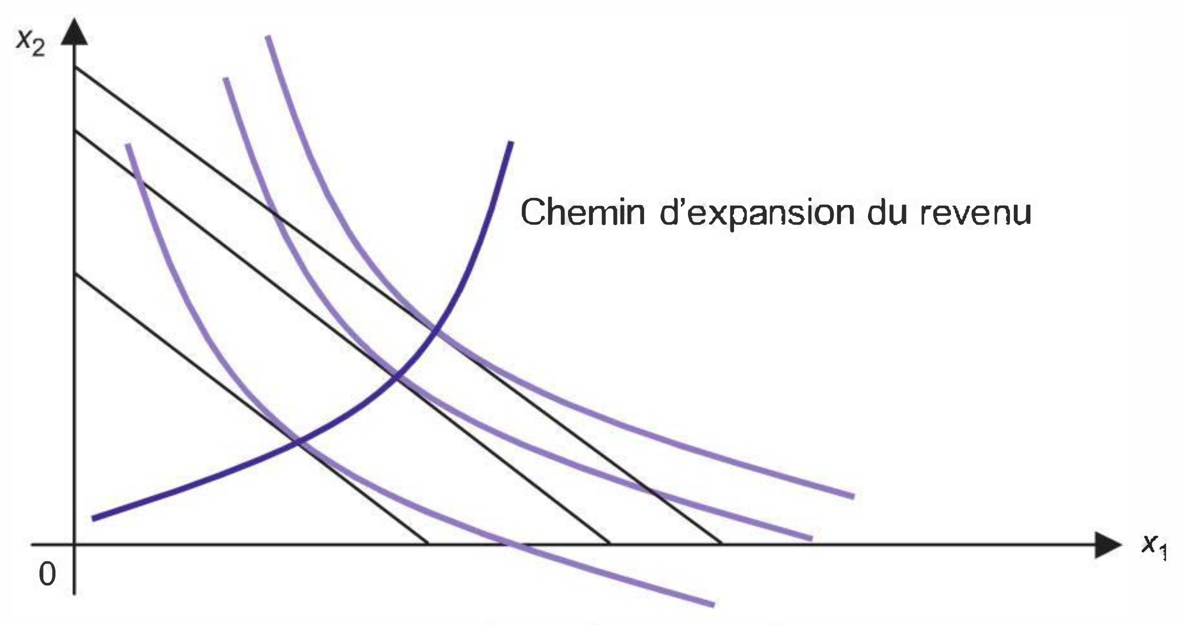


Figure 16 Chemin d'expansion du revenu

3.2 Courbe d'Engel

La consommation optimale x_i^* du bien i en fonction des différents niveaux du revenu R figurée dans le repère $(0\,;R\,;x_i)$ est appelée **courbe d'Engel**.

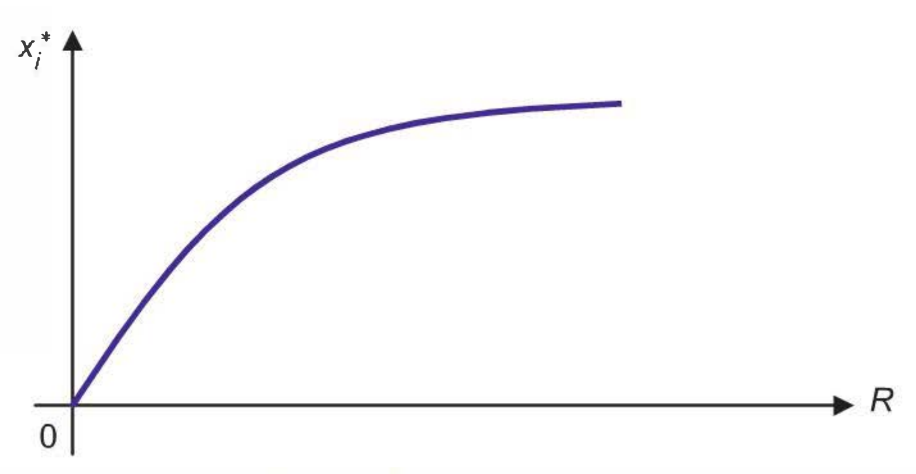


Figure 17 Courbe d'Engel

La courbe d'Engel est-elle nécessairement croissante ? Il se peut très bien qu'à la suite d'un accroissement de ses ressources, un consommateur décide de consommer moins d'un bien au profit d'un autre de meilleure

qualité ou de niveau de gamme plus élevé. Par exemple, la consommation d'aliments bio, fermiers ou de production artisanale pourra être privilégiée au détriment de produits issus de l'industrie agro-alimentaire traditionnelle, en particulier les aliments « premiers prix », généralement de plus faible qualité gustative, plus gras, ou subissant des adjonctions de colorants et d'arômes artificiels. Ces aliments ou biens de qualité moindre sont qualifiés de **biens inférieurs**. Au sens strict, des biens inférieurs sont des biens dont la consommation baisse dès que le revenu augmente. De manière plus réaliste, il s'agirait de biens dont la consommation varie en sens inverse du revenu, *à partir d'un certain seuil de revenu* (figure 18).

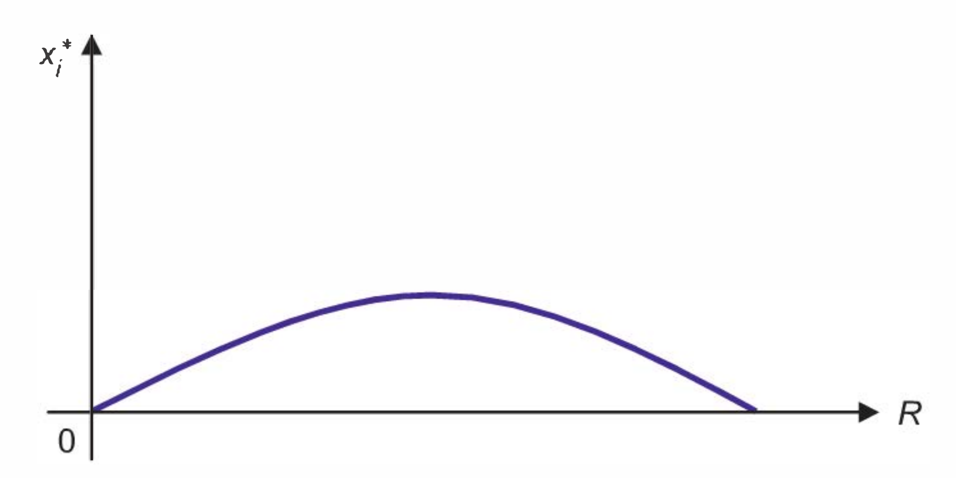


Figure 18 Courbe d'Engel relative à un bien de faible qualité

À l'opposé, certains biens ou services voient leur consommation augmenter plus que proportionnellement avec le revenu. C'est le cas de certains biens culturels (tableaux, sculptures, pièces de théâtre, opéras, etc.), mais aussi de certains vêtements de marques ou parfums, ainsi que des vins fins, tels que les Pessac Léognan ou les Saint Julien. Naturellement, ces biens sont appelés **biens de luxe**. En réalité, cette appellation est trompeuse car de nombreux biens de cette catégorie sont des biens auxquels il ne nous viendrait pas l'idée d'accoler le qualificatif luxueux. Il en va ainsi des dépenses de santé ou des dépenses relatives aux loisirs (vacances, sport, consoles de jeux, home cinéma, etc.), auxquelles les ménages consacrent parfois toutes les tranches de revenu supplémentaires au-delà d'un certain seuil de ressources (figure 19).

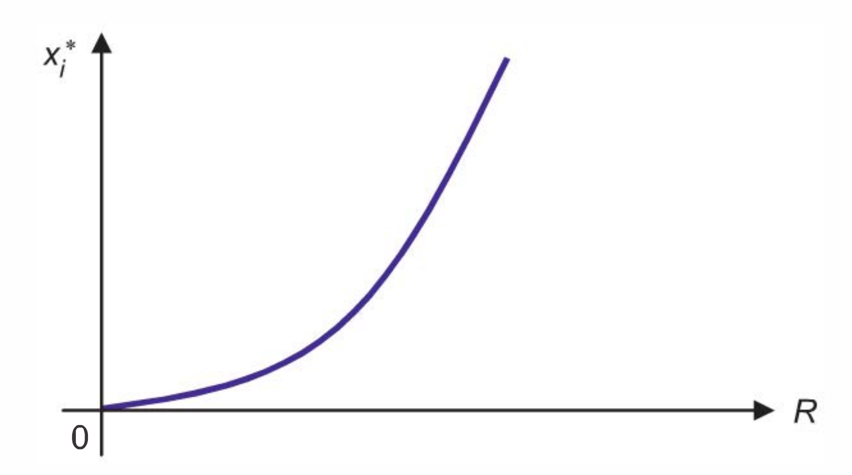


Figure 19 Courbe d'Engel relative à un bien de « luxe »

Les biens dits de luxe sont traditionnellement rangés dans la catégorie des biens dits **normaux** (par opposition aux biens inférieurs). Ils côtoient dans cette catégorie les biens dits **de nécessité**, dont la consommation, si elle croît avec le revenu, progresse de manière moins que proportionnelle. Traditionnellement, on considère que les biens alimentaires et vestimentaires communs (ni « premiers prix », ni de grandes marques) émargent au rang des biens de nécessité.

La manière la plus appropriée de caractériser les biens est de les classer en fonction de l'élasticité-revenu de leur demande.

3.3 Élasticité-revenu

On appelle **élasticité-revenu** de la demande d'un bien *i*, et l'on note $\varepsilon_{i;R}$, la mesure de la sensibilité de la demande de ce bien à un accroissement du revenu. On peut interpréter $\varepsilon_{i;R}$, comme le pourcentage d'augmentation (ou de baisse) de la demande en bien *i* induit par un accroissement de 1 % du revenu.

Typologie

- Un bien **inférieur** est un bien dont l'élasticité-revenu de la demande est négative.
- Un bien **normal** est un bien dont l'élasticité-revenu de la demande est positive :

- un bien **de nécessité** est un bien dont l'élasticité est comprise entre 0 et 1 ;
- un bien **de luxe** est un bien dont l'élasticité est supérieure à 1.

Pour aller plus loin

Élasticité

Une élasticité est une mesure normalisée de sensibilité d'une fonction relativement à l'une de ses variables. On la construit comme un quotient où figure, au numérateur, la dérivée partielle de la fonction relativement à la variable, et, au dénominateur, le rapport des valeurs respectives de la fonction et de la variable. Ainsi, dans le cas d'une fonction de demande $x_i(p_1, p_2, \ldots, p_n, R)$, l'élasticité-revenu de la demande est :

$$\varepsilon_{i;R} = \frac{\dfrac{\partial x_i(p_1, p_2, \ldots, p_n, R)}{\partial R}}{\dfrac{x_i(p_1, p_2, \ldots, p_n, R)}{R}} = \frac{\partial x_i(p_1, p_2, \ldots, p_n, R)}{\partial R} \times \frac{R}{x_i(p_1, p_2, \ldots, p_n, R)}$$

4 Impact d'une variation de prix

4.1 Effets substitution et revenu

Suite à l'accroissement du prix de l'un des biens (bien 1 dans l'exemple illustré sur la figure 20), l'ensemble budgétaire se contracte. Il est facile de voir comment la décision optimale de consommation se modifie (passage du panier A au panier C).

Les niveaux de satisfaction atteints, à l'optimum, sont U_A avant l'augmentation du prix du bien 1 et U_C après l'augmentation.

Dans le cas présenté ici, l'augmentation du prix du bien 1 a, sans équivoque, fait baisser la demande en bien 1. Elle a aussi conduit à un accroissement de la demande en bien 2. Autant la diminution de la demande de bien 1 semble un phénomène franc et peu dépendant des caractéristiques du faisceau de courbes d'indifférence, autant l'accroissement de la demande de bien 2 semble un phénomène fragile et dépendant des caractéristiques

des préférences du décideur. En effet, la demande exprimée pour le bien dont le prix ne s'est pas modifié parfois diminuera, à l'inverse, et parfois même sera inchangée.

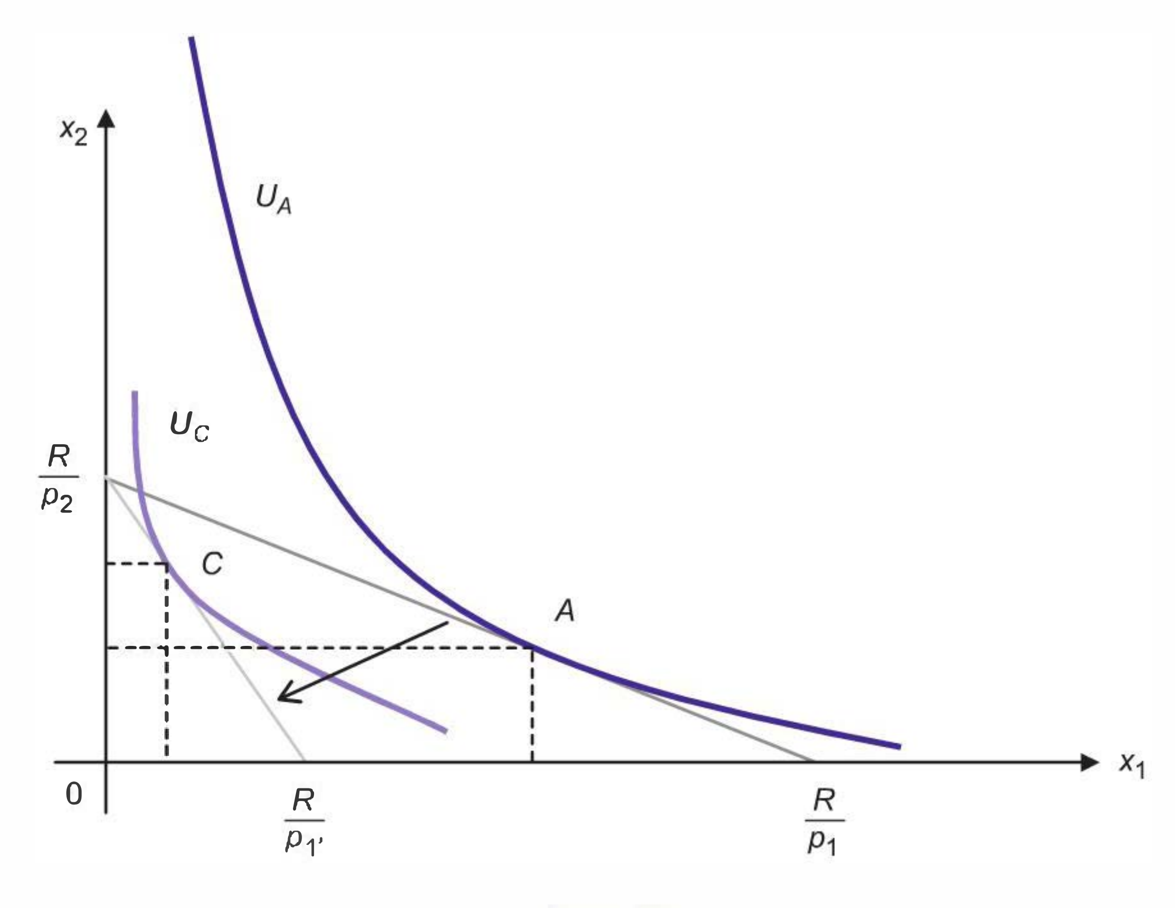


Figure 20

Quelles sont les forces économiques en jeu ? On peut très naturellement identifier deux forces antagonistes ayant un rôle quant à l'évolution de la demande *du bien dont le prix ne s'est pas modifié* (ici le bien 2) :

- Il y a tout d'abord la tentation de consommer plus de ce bien dans la mesure où il est devenu relativement moins cher. Comme nous avons commencé à le comprendre dans ce chapitre, ce qui détermine le comportement du consommateur, c'est l'observation des prix relatifs (les prix absolus n'ont pas de signification en soi). Le prix du bien 2 n'a pas changé, mais comme celui du bien 1 a augmenté, le bien 2 est devenu relativement moins cher, ce qui pousse le consommateur à substituer de la consommation de bien 1 par de la consommation de bien 2. Pour cette raison, on parle d'**effet substitution**.

- Il y a aussi la tentation de consommer moins de ce bien (comme de tous les autres) car la capacité d'achat générale du consommateur a décru dans la mesure où le revenu est resté identique alors que le prix de l'un des biens progressait (on peut donc parler de baisse du revenu relatif). Il s'agit d'un effet de dégradation du pouvoir d'achat que l'on a pour coutume de nommer **effet revenu**.

Ces deux effets peuvent aussi être mis en relief lorsque l'on cherche à analyser la variation de la demande de bien 1. Mais, dans ce cas précis, les deux effets se cumulent dans le sens d'une baisse de la demande.

L'effet substitution est l'effet, sur la demande des biens, de la modification des prix relatifs des biens, qui conduit le consommateur à consommer moins des biens devenus relativement plus chers et plus des biens devenus relativement moins chers.

L'effet revenu est l'effet, sur la demande des biens, de la variation de pouvoir d'achat induite par la variation du revenu relatif, qui pousse le consommateur à consommer moins de chacun des biens lorsque le revenu relatif diminue, et plus de chacun des biens lorsque le revenu relatif augmente.

Il serait intéressant de pouvoir quantifier plus précisément ces deux effets. L'observation du seul graphique ci-dessus ne le permet pas. Il nous faut avoir de l'imagination et concevoir un dispositif permettant d'isoler l'un des deux effets.

Essayons de neutraliser l'effet revenu ; il deviendra alors facile de mettre en relief l'effet substitution. Une telle neutralisation de l'effet revenu ne correspond pas à ce qui se passe effectivement dans la réalité : il s'agit bien d'un dispositif artificiel. En quoi consiste-t-il ? Nous allons supposer qu'est attribué au consommateur un supplément de revenu ΔR, précisément calculé pour qu'il lui permette de conserver le niveau de satisfaction obtenu avant l'augmentation du prix du bien (c'est-à-dire U_A). Un tel supplément de revenu est appelé **variation compensatrice de revenu**.

> Une **variation compensatrice de revenu** est un supplément (ou une baisse) artificiel de revenu accordé au consommateur lui permettant de maintenir sa satisfaction au niveau initial.

Dans cette configuration artificielle où, en plus de subir l'accroissement du prix du bien 1 de p_1 à p_1', le consommateur se verrait désormais doté d'un revenu égal à $R + \Delta R$, celui-ci atteindrait l'optimum B (figure 21).

Figure 21

Le panier B est l'optimum du consommateur lorsque les prix des biens sont respectivement p_1' et p_2 et que son revenu est $R + \Delta R$. La droite de budget est de pente $-\frac{p_1'}{p_2}$, elle est donc parallèle à la droite de budget conduisant à l'optimum C. Au point B, elle est nécessairement tangente à la courbe d'indifférence U_A (c'est ainsi que l'on peut, graphiquement établir l'intensité

de la variation compensatrice de revenu ΔR). Puisque le passage de l'optimum A à l'optimum B correspond à l'impact de l'accroissement du prix du bien 1 *une fois neutralisé l'effet revenu*, on considérera que le passage de A à B est caractéristique de l'effet substitution. Par différence, le passage de B à C sera réputé caractéristique de l'effet revenu. Résumons ces différents effets dans un tableau :

Augmentation de p_1	Effet sur la demande de bien 1	Effet sur la demande de bien 2
Effet substitution	–	+
Effet revenu	–	–
Effet total	–	Indéterminé

Dans le cas présent, la demande en bien 2 a augmenté suite à l'accroissement du prix du bien 1 : l'effet total est donc positif. Cela suggère que, pour la demande de bien 2, l'effet substitution a dominé l'effet revenu. Mais il se pourrait que, suite à un nouvel accroissement du prix du bien 1, l'effet revenu (négatif) vienne à dominer l'effet substitution. Dès lors, la demande de bien 2 diminuerait. On peut résumer ces situations en traçant une courbe qui indique comment la demande de bien 2 évolue suite à un accroissement du prix du bien 1 (figure 22).

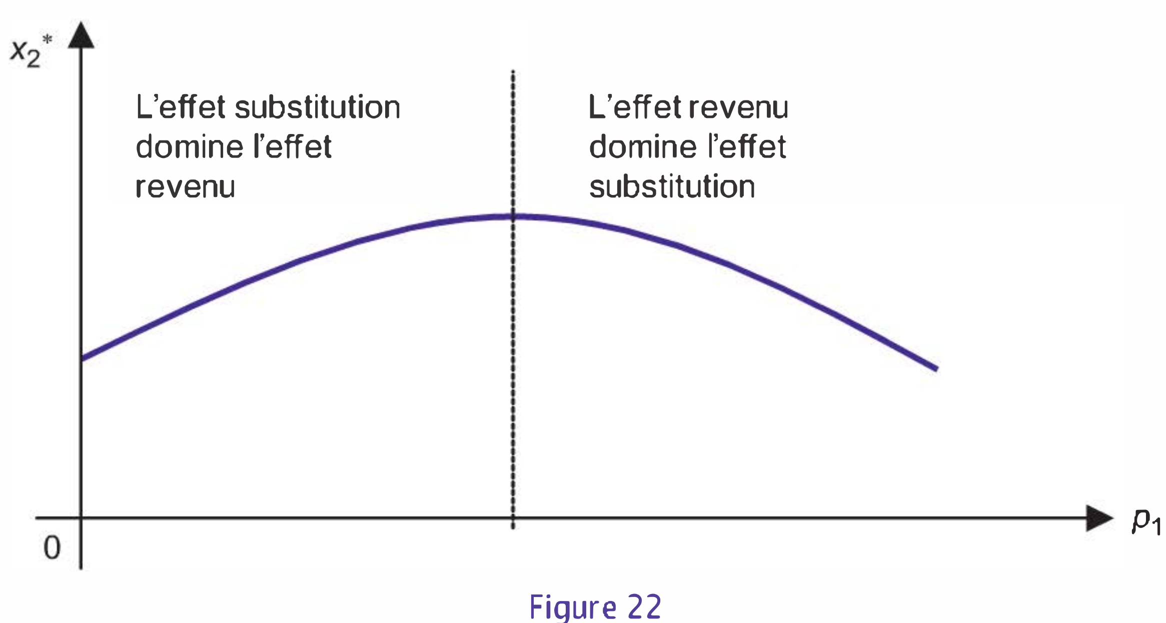


Figure 22

4.2 Élasticités-prix

■ Élasticité-prix croisée

Une autre manière de caractériser l'évolution de la demande en bien *i* suite aux variations du prix d'un bien *j* est de calculer l'élasticité-prix croisée.

On appelle **élasticité-prix croisée de la demande d'un bien *i* relativement au prix du bien *j*,** et l'on note $\varepsilon_{i\,;\,j}$, la mesure de la sensibilité de la demande du bien *i* à un accroissement du prix du bien *j*. On peut interpréter $\varepsilon_{i\,;\,j}$, comme le pourcentage d'augmentation (ou de baisse) de la demande en bien *i* induit par un accroissement de 1 % du prix du bien *j*.

■ Élasticité-prix directe

Quelles sont les conséquences d'un accroissement du prix du bien *i* sur la demande de ce même bien *i* ? Presque toujours, tout accroissement du prix d'un bien conduit à une baisse de la demande de ce bien car les effets substitution et revenu se cumulent pour la faire baisser. Néanmoins, il faut envisager des cas dans lesquels les choses ne se passent pas ainsi.

Se peut-il que la demande d'un bien croisse quand son prix augmente ? Le premier cas historiquement rapporté d'un tel phénomène le fut, semble-t-il, par l'économiste écossais Giffen. En effet, ce dernier eut l'opportunité d'observer, à l'occasion d'une situation de pénurie alimentaire en Irlande, que les individus consommaient de plus en plus de pommes de terre alors que le prix desdites pommes de terre augmentait. On qualifie depuis de **biens de Giffen** les biens dont la consommation augmente lorsque leur prix augmente.

Pour quel raison un tel phénomène, *a priori* contre-intuitif, a-t-il pu se produire ? Les ménages irlandais étaient confrontés à une augmentation de l'ensemble des prix alimentaires à l'occasion de cette pénurie. S'ils choisissaient de consacrer plus de leurs ressources à l'acquisition de pommes de terre (et considérablement moins à l'acquisition d'autres biens) c'est en raison de la valeur calorique des pommes de terre : dans des situations extrêmes, la question n'est plus celle du goût pour la variété (viande, produits laitiers, fruits, légumes, féculents, etc.), mais bien celle de

la nécessité absolue de se procurer des aliments riches en calories pour survivre ! Le phénomène observé par Giffen est très singulier et s'inscrit dans un cadre à peine conforme aux hypothèses de travail retenues jusqu'ici. Néanmoins, l'idée que la demande d'un bien puisse croître avec son prix ne doit pas être totalement écartée.

Il existe tout d'abord un bien dont il est théoriquement possible de voir la demande croître suite à une augmentation de son prix : ce bien, c'est la *couverture assurantielle* ou *assurance*. En effet, si, en cas d'augmentation de son prix, l'effet de substitution conduit bien à faire diminuer la demande de ce bien, le sens de l'effet revenu est susceptible d'être inversé. Plutôt que d'effet revenu, on parle d'ailleurs d'effet richesse : la baisse de pouvoir d'achat peut entraîner un accroissement de l'aversion pour le risque tel que la demande de couverture assurantielle augmente au titre de cet effet. Ainsi, si l'effet richesse domine l'effet de substitution, la demande d'assurance peut globalement croître suite à l'accroissement de son prix.

Il existe ensuite des biens très particuliers dont la demande peut croître avec le prix : ce sont les biens ostentatoires, ceux dont la possession permet à leurs propriétaires de faire connaître le niveau de leur fortune ou d'affirmer leur appartenance à une classe dirigeante. Le prix des Ferrari ne cesse de progresser, mais il y a toujours plus d'acheteurs potentiels ; il en va de même pour les bouteilles de Château Haut Brion ou pour certains sacs produits par le maroquinier Louis Vuitton. On parle aussi de **biens de Veblen** (du nom d'un économiste américain d'origine norvégienne) pour désigner ce type de biens.

Les biens de Giffen et de Veblen ont donc en commun de voir leur demande croître avec leur prix. Ceci est généralement étudié à partir de l'observation de l'élasticité-prix directe de leur demande.

On appelle **élasticité-prix directe de la demande d'un bien *i* (relativement au prix du bien *i*)**, et l'on note $\varepsilon_{i;i}$, la mesure de la sensibilité de la demande du bien *i* à un accroissement de son prix. On peut interpréter $\varepsilon_{i;i}$, comme le pourcentage d'augmentation (ou de baisse) de la demande en bien *i* induit par un accroissement de 1 % de son prix.

Typologie

- Un bien **ordinaire** est un bien dont l'élasticité-prix directe est négative.
- Un bien **de Giffen** est un bien dont l'élasticité-prix directe est positive.

Pour clore ce chapitre, il convient de mentionner l'existence de biens ayant une propriété fascinante : celle de modifier les préférences de l'individu au fur et à mesure que celui-ci consomme de ces biens. En particulier, le fait de consommer certains biens accroît l'appétence du consommateur pour ces biens. Il peut en aller ainsi pour la drogue, les jeux de casino, la musique baroque, les vins de Bordeaux… et même les concours de beauté canine (bien que cela demeure en partie inexplicable). On parle de **biens addictifs**. La présence d'un bien addictif dans un panier peut conduire à de la concavité des préférences.

Pour aller plus loin

Élasticité-prix directe et élasticité-prix croisée de la demande

Soit une fonction de demande $x_i(p_1, p_2, \ldots, p_n, R)$.
On définit l'élasticité-prix directe de la demande en bien i, $\varepsilon_{i\,;\,i}$ par :

$$\varepsilon_{i;i} = \frac{\dfrac{\partial x_i(p_1, p_2, \ldots, p_n, R)}{\partial p_i}}{\dfrac{x_i(p_1, p_2, \ldots, p_n, R)}{p_i}} = \frac{\partial x_i(p_1, p_2, \ldots, p_n, R)}{\partial p_i} \times \frac{p_i}{x_i(p_1, p_2, \ldots, p_n, R)}$$

On définit l'élasticité-prix croisée de la demande en bien i relativement au prix du bien j, $\varepsilon_{i\,;\,j}$ par :

$$\varepsilon_{i;j} = \frac{\dfrac{\partial x_i(p_1, p_2, \ldots, p_n, R)}{\partial p_j}}{\dfrac{x_i(p_1, p_2, \ldots, p_n, R)}{p_j}} = \frac{\partial x_i(p_1, p_2, \ldots, p_n, R)}{\partial p_j} \times \frac{p_j}{x_i(p_1, p_2, \ldots, p_n, R)}$$

3 Travail et loisir, horizon temporel, incertitude

Mots-clés

Arbitrage consommation-loisir, arbitrage intertemporel de consommation, choix d'un panier de biens contingents, assurance.

Les biens entrant dans la composition d'un panier peuvent être des biens et services au sens usuel du terme : biens alimentaires, vêtements, smartphones, meubles, logements, véhicules, carburant, fourniture d'eau, de gaz, d'éléctricité, d'accès internet, séance de cinéma, match de football, trajet ferroviaire, prestation d'artisan plombier, prestation de nettoyage, etc. On peut aussi envisager d'autres biens plus abstraits à l'image du bien « loisir » défini comme le temps que le consommateur ne consacre pas au travail salarié. En outre, un panier peut contenir des biens consommés à différentes dates ou des biens dits « contingents », c'est-à-dire dont la quantité consommée dépend de la survenance d'un aléa. Ce sont tous ces paniers de biens au sens élargi du terme dont nous étudions la consommation dans ce chapitre.

1 L'arbitrage consommation-loisir

Pour rester dans un cadre simple, nous supposons que le panier de biens et services usuels sur lequel le consommateur construit sa décision peut être « résumé » en un unique bien de consommation. Cette synthèse de sa consommation de biens traditionnels est désignée par l'appellation « bien composite ».

On appelle **bien composite** un bien virtuel dont la quantité consommée incarnerait les quantités consommées des différents biens et services demandés par le consommateur.

À l'aide de cette notion de bien composite, il est possible de rester dans le cadre très simple où $n = 2$ biens pour étudier l'arbitrage consommation-loisir. Le second bien susceptible d'être acquis par le consommateur est en effet le bien « loisir ». Le loisir est, par définition, le temps « non consacré » au travail salarié. Il s'agit donc d'un gigantesque « fourre-tout » contenant, outre le loisir au sens usuel (cinéma, lecture, sport, etc.) tous les moments de la vie non utilisés pour l'activité salariée : études supérieures, sommeil, tâches ménagères, repas, ablutions, formalités administratives, etc. Une fois admis qu'il est possible de raisonner sur le loisir comme sur n'importe quel bien de consommation désiré par le consommateur, le modélisateur est néanmoins confronté à deux questions : quel est le prix du bien « loisir » et les ressources du consommateur demeurent-elles indépendantes de sa décision de consommation ? Tentons d'abord de répondre à la seconde question. Plus le consommateur travaille, moins il « demande » de loisir, mais plus ses ressources salariales sont élevées. À l'inverse, moins il travaille, plus il demande de loisir, et moins ses ressources salariales sont importantes. On comprend donc que le niveau de ses ressources est impacté par son choix de consommation du bien « loisir », choix lui-même dépendant du niveau de ses ressources ! Ainsi ses ressources et le niveau de sa consommation sont interdépendantes ce qui est une différence notable avec le cas « standard » étudié dans le chapitre 2. Pour répondre à la question sur le prix du bien « loisir », nous allons devoir entrer dans la formalisation du problème.

Raisonnons sur une période de temps donnée, notée ℓ_0 (par exemple $\ell_0 = 24$ heures ou 7 jours ou 1 mois ...). À l'intérieur de cette période, l'individu choisit de travailler une certaine durée ℓ ($\ell \leq \ell_0$). Pour ce travail, l'individu est rémunéré au taux w (taux de salaire), soit, en niveau $w \times \ell$. Ce revenu salarial peut, ou non, être complété par un revenu non salarial noté R (royalties, droits d'auteurs, intérêts de placements financiers, loyers perçus, etc.). Les ressources monétaires totales du consommateur sont donc $w \times \ell + R$. En termes de décaissements « physiques » d'argent, le seul bien

consommé est l'unique bien composite acheté au prix unitaire p. Nous notons c la quantité consommée de ce bien. Il est désormais possible d'écrire, sous une première forme, la contrainte budgétaire du consommateur. Cette première forme correspond aux encaissements et décaissements réels (ou physiques) d'argent :

$$pc \leq w\ell + R$$

En ajoutant le terme $w\ell_0$ aux deux membres de cette inégalité, on obtient :

$$pc + w\ell_0 \leq w\ell + R + w\ell_0$$

ou encore :

$$pc + w(\ell_0 - \ell) \leq w\ell_0 + R$$

Cette seconde écriture intègre un décaissement et un encaissement virtuel : $w(\ell_0 - \ell)$ mesure la dépense d'acquisition du bien « loisir », qui est un décaissement virtuel ; quant au terme $w\ell_0$, il désigne les ressources salariales maximales que le consommateur obtiendrait s'il décidait de consacrer l'intégralité de son temps disponible au travail salarié. Puisqu'il ne travaille pas 24 heures sur 24, une partie de ces ressources potentielles ne sont pas obtenues. Cette partie non obtenue des ressources potentielles constitue un encaissement virtuel.

En écrivant sous cette seconde forme la contrainte budgétaire du consommateur, nous faisons apparaître le prix du bien « loisir ». Dans le terme de gauche de l'inégalité (du côté des dépenses), la quantité consommée de bien loisir est ainsi valorisée au prix w. Ce prix du loisir est un coût d'opportunité.

On appelle **coût d'opportunité** d'un bien ou service le prix de la renonciation à la jouissance de ce bien ou service. En d'autres termes, c'est le coût du bien évalué en termes d'opportunités non concrétisées.

Ce concept est complexe mais central dans la théorie économique. Dans le cas du loisir, tout se passe comme si l'individu « achetait » du loisir au prix w. Chaque heure de loisir correspond à un coût égal aux ressources salariales dont l'individu se prive du fait du temps consacré aux loisirs. C'est

un coût d'opportunité puisqu'il s'agit d'une perte de ressources potentielles due au fait que l'individu renonce à travailler pendant l'intégralité de la dotation en temps disponible (mais pendant une fraction seulement).

Comme précédemment, le consommateur tente de prendre la meilleure décision possible en maximisant sa satisfaction sous la contrainte « budgétaire » exprimée ci-dessus. Cela signifie qu'il est possible de définir une fonction d'utilité (à partir des préférences de l'agent) sur un panier composé du bien composite et du bien « loisir ». Nous notons $u[c; (\ell_0 - \ell)]$ cette fonction. Cette fonction est, bien sûr, supposée croissante en ses deux arguments : la satisfaction croît avec la quantité consommée de bien composite d'une part, avec la quantité consommée de loisir d'autre part.

Figurons maintenant l'ensemble budgétaire du consommateur dans un repère construit avec le nombre d'unités consommées de bien composite en abscisse et le nombre d'heures (si ℓ_0 est exprimée en heures) consommées de loisir. Outre la contrainte budgétaire exprimée ci-dessus, nous devons faire apparaître une seconde contrainte qui indique que le nombre d'heures de loisirs consommées ne doit pas excéder la dotation totale en temps disponible (je ne peux pas consommer plus de 24 heures de loisir par jour !). Cette contrainte est de la forme :

$$(\ell_0 - \ell) \leq \ell_0$$

La présence conjointe des deux contraintes va donner à l'ensemble budgétaire une forme de trapèze, semblable à celle que l'on retrouve en présence d'un rationnement (cf. chapitre 2). Faisons apparaître, sur la figure 1, l'ensemble budgétaire du consommateur.

Comme précédemment, la décision optimale est obtenue pour la plus haute courbe d'indifférence compatible avec l'ensemble budgétaire. Deux configurations sont possibles, selon le niveau du revenu non salarial R :

- soit le revenu non salarial R est faible, auquel cas le consommateur ne bute pas sur le désir de consommer une quantité de loisir supérieure à ℓ_0 ; dans ce cas, le consommateur doit effectivement consacrer un certain nombre d'heures à offrir du travail salarié et sa décision optimale de consommation est telle que sa courbe d'indifférence est tangente à la contrainte budgétaire (figure 2).

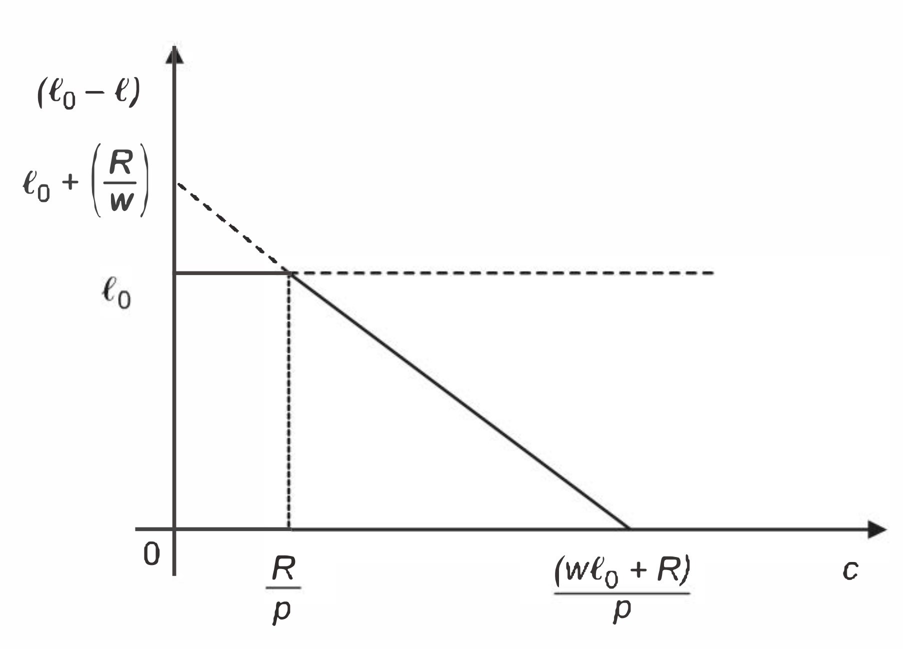


Figure 1

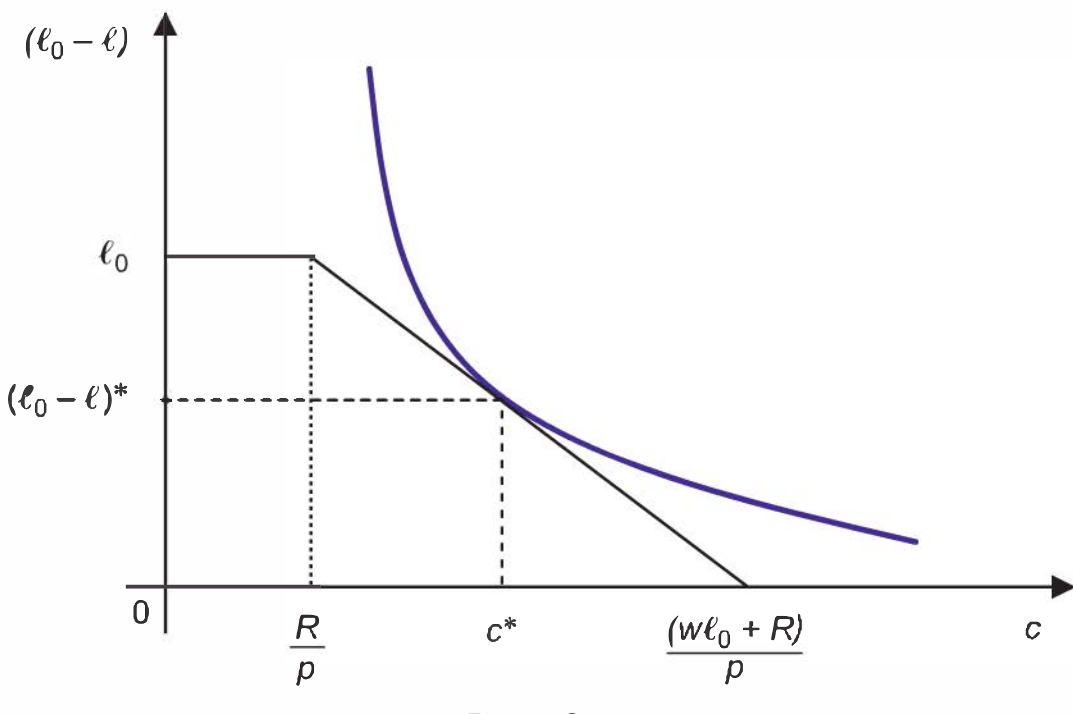


Figure 2

- soit le revenu non salarial R est très élevé, auquel cas le consommateur peut s'abstenir de travailler (s'agissant du travail salarié). En d'autres termes, il bute sur la contrainte l'empêchant de consommer une quantité de loisir supérieure à ℓ_0 (il aimerait que chacune de ses journées dure 48 heures, tant il a de choses exaltantes à faire !) (figure 3).

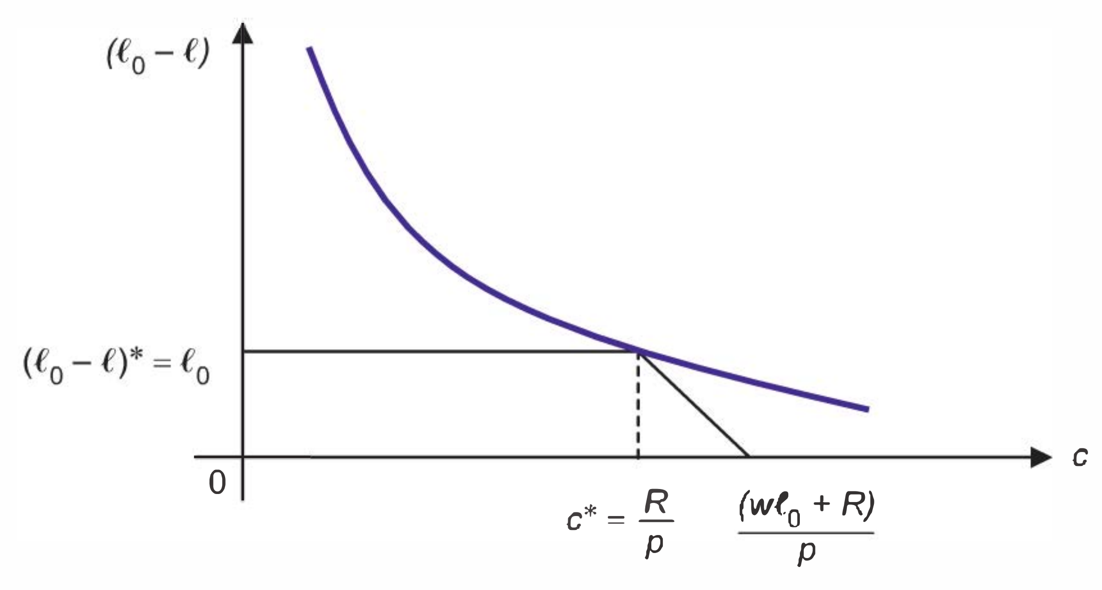


Figure 3

Pour aller plus loin

Résolution du programme du consommateur

Le problème que nous avons résolu graphiquement ci-dessus peut être posé de manière formelle. La recherche du plus haut niveau de satisfaction possible correspond, comme précédemment, à la maximisation de la fonction d'utilité du consommateur (et le fait que l'utilité éprouvée dépende de la consommation de loisir ne modifie rien à la démarche). La maximisation se fait sous la contrainte budgétaire présentée et sous la seconde contrainte qui indique que le nombre d'heures consommées de loisirs ne doit pas excéder la dotation totale en temps disponible. Ainsi, le consommateur doit résoudre le programme suivant :

$$\begin{aligned} & \underset{c,(\ell_0-\ell)}{Max} \quad u[c\,;(\ell_0-\ell)] \\ s.c. \quad & pc + w(\ell_0-\ell) \leq w\ell_0 \\ & (\ell_0-\ell) \leq \ell_0 \end{aligned}$$

Comme précédemment, le consommateur bute toujours sur sa contrainte « budgétaire » (en raison du respect de l'axiome de non-saturation). En revanche, la seconde contrainte n'est pas nécessairement saturée. Il faut dès lors, en fonction du niveau de R, distinguer deux cas :

- Soit la seconde contrainte n'est pas saturée, auquel cas la décision optimale $[c^*\,;(\ell_0-\ell)^*]$ est caractérisée par $TmS_{(\ell_0-\ell)/c} = -\dfrac{\dfrac{\partial u}{\partial(\ell_0-\ell)}}{\dfrac{\partial u}{\partial c}} = -\dfrac{w}{p}$.

Ceci peut se réécrire $\dfrac{\dfrac{\partial u}{\partial(\ell_0 - \ell)}}{w} = \dfrac{\dfrac{\partial u}{\partial c}}{p}$, ou encore $\dfrac{Um_{(\ell_0-\ell)}}{w} = \dfrac{Um_c}{p}$ (condition équi-marginale) : le surcroît de satisfaction par euro dépensé obtenu grâce à la dernière unité consommée de bien loisir est égal au surcroît de satisfaction par euro dépensé obtenu grâce à la dernière unité consommée de bien composite.

- Soit la seconde contrainte est saturée, auquel cas la condition équi-marginale n'est pas vérifiée et la décision optimale est $c^* = \dfrac{R}{p}$ et $(\ell_0 - \ell)^* = \ell_0$.

Une application intéressante de l'arbitrage consommation-loisir est l'analyse de l'impact d'un accroissement du salaire. Le salaire w étant à la fois la rémunération de l'heure de travail et le prix unitaire du bien loisir, on devine que plusieurs effets antagonistes émergent lorsque celui-ci augmente : d'une part, un effet associé à la volonté du consommateur-salarié d'offrir plus de travail (c'est-à-dire de travailler plus) car celui-ci est plus récompensé, et donc de demander corrélativement moins de loisir ; d'autre part, le souhait de consommer plus de bien loisir car l'augmentation du revenu salarial pousse à consommer plus de chacun des biens ; enfin, la volonté de consommer moins du bien loisir devenu relativement plus cher que le bien composite. Nous retrouvons l'intuition d'un effet substitution comme dans le chapitre 2 et nous identifions également une forme d'effet revenu ; nous voyons de plus émerger un effet mécanique de baisse de la demande de loisir corrélatif à la hausse de l'offre de travail. L'effet revenu n'est pas exactement comparable à celui identifié précédemment : cette fois-ci, le revenu *absolu* change suite à la variation de w, tandis qu'auparavant seul le revenu *relatif* était altéré. Quant au nouvel effet, il n'est pas aisé de le dissocier pleinement des deux précédents. Pour simplifier, nous prétendrons que coexistent deux effets : un **effet substitution** et un effet **richesse** (pour le différencier de l'effet revenu précédemment invoqué).

L'effet richesse est l'effet, sur la demande des biens, de la variation des ressources du consommateur, qui le pousse à consommer moins de chacun des biens lorsque ces ressources diminuent, et plus de chacun des biens lorsque ces ressources augmentent.

Le tableau suivant résume les différents effets induits par un accroissement de w :

Augmentation de w	Demande de bien composite	Demande de loisir
Effet substitution	+	–
Effet richesse	+	+
Effet total	+	Indéterminé

Ainsi l'effet total de l'accroissement de w sur la demande de loisir est indéterminé. On peut envisager deux situations :

- Tout d'abord, suite à un premier accroissement du salaire (on suppose par exemple que le salaire passe de w à w', avec $w' > w$), la demande de loisir diminue. Ceci est la conséquence de ce que l'effet substitution domine l'effet richesse. Le consommateur accroît prioritairement son offre de travail salarié pour bénéficier de l'accroissement de la rémunération horaire du travail. Ainsi, le bien « loisir » est-il un bien ordinaire, bien dont la demande baisse lorsque son prix s'accroît.
- Puis, suite à un second accroissement du salaire (le salaire passe de w' à w'', avec $w'' > w'$), la demande de loisir augmente. Ceci est la conséquence de ce que l'effet richesse domine l'effet substitution. Le consommateur réduit désormais son offre de travail salarié pour bénéficier de l'opportunité qu'il a de consommer autant (de bien composite) en travaillant moins. Dès lors, le bien « loisir » s'apparente à un bien de Giffen, bien dont la demande augmente lorsque son prix s'accroît.

Illustrons cet impact d'un double accroissement du salaire w (le prix du bien composite p demeurant inchangé). Étudions, sur la figure 4, comment se déplace la contrainte budgétaire suite à un accroissement de salaire.

Puisque la pente de la droite (en valeur absolue) est p/w, la droite devient moins pentue. L'abscisse à l'origine et l'ordonnée à l'origine se modifient. Le seul point par lequel passe toutes les droites de budget est le point A d'abscisse R/p et d'ordonnée ℓ_0.

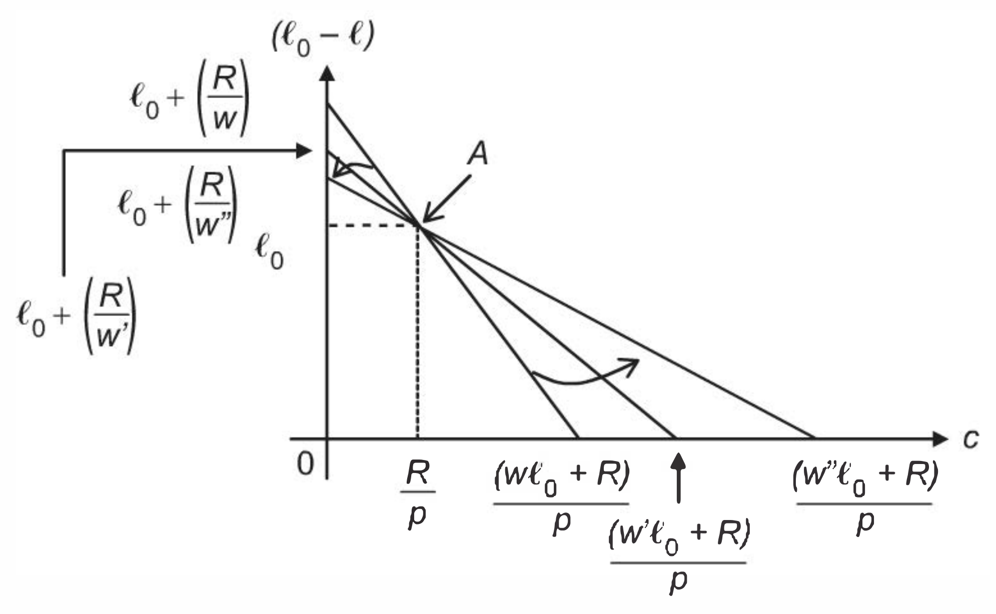


Figure 4

Faisons maintenant apparaître, sur la figure 5, la décision optimale du consommateur dans les 3 situations, en positionnant les plus hautes courbes d'indifférence compatibles avec chacun des ensembles budgétaires.

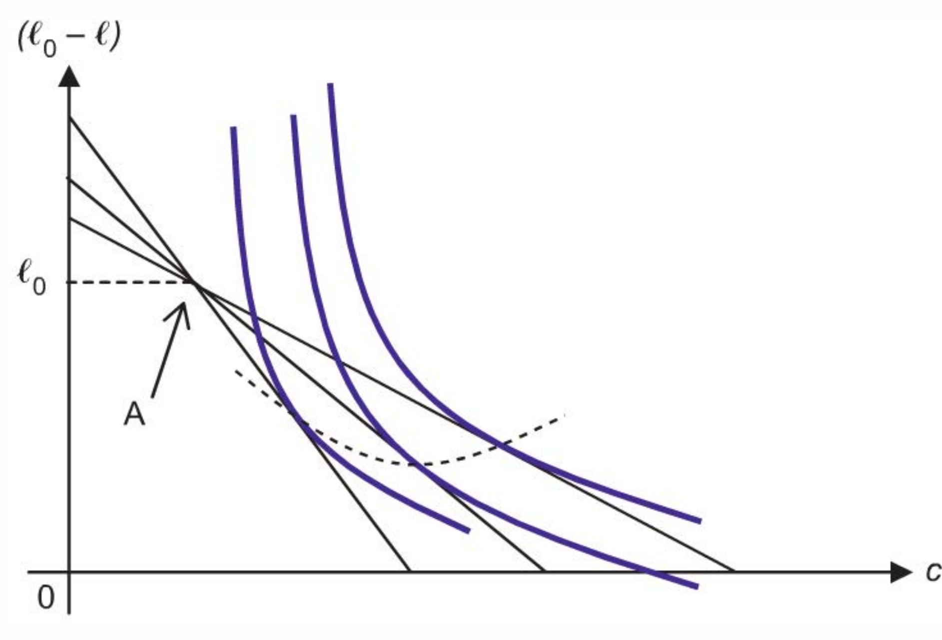


Figure 5

Il y a donc un seuil de salaire horaire en dessous duquel l'effet substitution domine l'effet richesse et au-dessus duquel l'effet richesse domine l'effet substitution. En conséquence le profil de la demande de loisir en fonction du salaire réel (le salaire divisé par le prix du bien composite) est le suivant (figure 6).

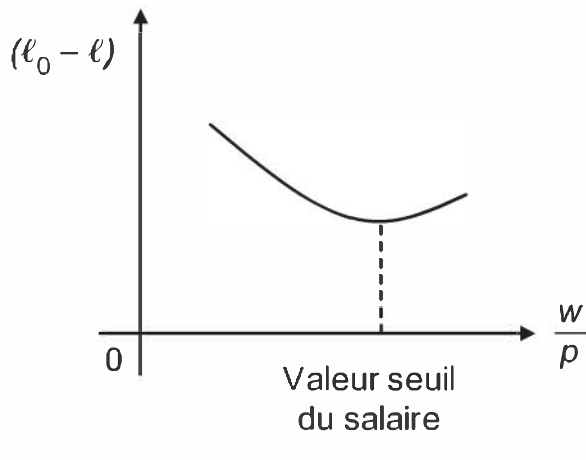


Figure 6

Une conséquente intéressante de ce phénomène est que l'offre de travail salarié ℓ n'est pas une fonction nécessairement croissante du salaire réel. En effet, par contraste, le profil de l'offre de travail salarié en fonction du salaire réel est le suivant (figure 7).

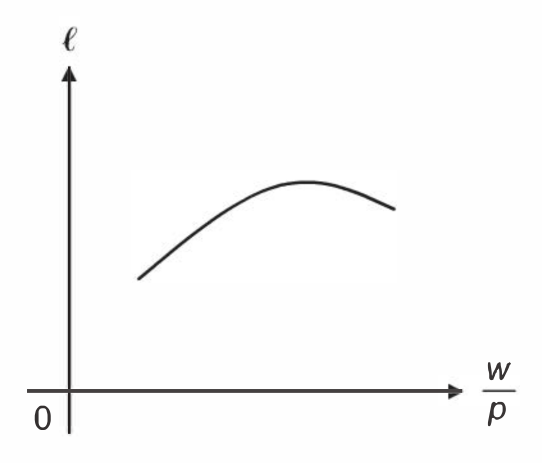


Figure 7

L'offre de travail est une fonction d'abord croissante, puis décroissante du salaire réel. Ce résultat peut avoir des conséquences majeures en macroéconomie. Une telle application démontre que la recherche de fondements microéconomiques aux comportements macroéconomiques est indispensable.

2 L'arbitrage intertemporel de consommation

Les consommateurs vivent plusieurs « périodes » de temps, plusieurs années ou plusieurs « tranches » de vie. Lorsqu'ils prennent des décisions de consommation pour une période particulière, ils essayent de tenir compte de ce qu'il se passe lors des autres périodes.

Par exemple, un consommateur désireux de faire un voyage en Polynésie choisit de ne pas dépenser une partie des ressources gagnées cette année afin d'accumuler de quoi partir l'an prochain. Il en va de même de toutes les décisions d'épargne qui permettent de différer l'utilisation de ressources obtenues lors de la période présente. À l'inverse, un consommateur désireux d'acquérir une maison ou un appartement, recourt à un emprunt qui lui permet d'anticiper l'utilisation de ressources obtenues lors de périodes futures. De même, un étudiant confronté à la nécessité de financer ses études (droits d'inscription, logement, dépenses courantes) doit parfois recourir à un emprunt qu'il remboursera à partir des salaires qu'il obtiendra plus tard, après ses études, lorsqu'il occupera un emploi salarié. Ce genre de mécanismes peut s'analyser simplement dans le cadre du modèle de décision du consommateur qui nous est désormais familier.

Il faut néanmoins supposer que la décision du consommateur n'est pas affectée par la présence d'incertitude : les flux de revenus futurs de l'agent sont supposés parfaitement connus ! Il s'agit bien sûr d'une hypothèse d'école, destinée à nous simplifier la tâche. La troisième section de ce chapitre sera consacrée à la prise en compte de l'incertitude et il est tout à fait possible de mêler la dimension temporelle et l'appréhension de l'incertitude dans une même modélisation. Mais pour dégager les mécanismes fondamentaux qui président à l'arbitrage intertemporel de consommation, nous ignorons, dans cette section, la dimension aléatoire du flux de ressources futures.

Le point central de l'analyse est la question du lissage de la consommation au fil du temps : qu'est ce qui explique que les consommateurs ont un profil de consommation mois après mois (ou année après année) plus régulier que celui de leurs ressources ? Intuitivement, on comprend que les

individus n'apprécient pas de baisser drastiquement leur « train de vie » quand leurs revenus fléchissent : il n'est jamais agréable de devoir renoncer aux dîners dans de bons restaurants ou aux escapades de fins de semaines sous prétexte de fins de mois difficiles. *A contrario*, une progression subite et durable des ressources conduit rarement un consommateur à changer brutalement de mode de vie. Dans les faits, les consommateurs semblent ajuster leur train de vie à ce qu'ils anticipent être le niveau moyen de leurs ressources, sur longue période. Comme nous allons le voir, ceci est le fruit d'un comportement rationnel si les préférences du consommateur sont convexes. Cela signifie donc que la convexité des préférences, non seulement synonyme de goût pour la variété des produits, est aussi la marque d'un goût pour la régularité, dans le temps, du « train de vie ».

Supposons que le consommateur dont nous étudions la décision ne « vive » que 2 périodes (pensons par exemple à deux périodes, d'une vingtaine d'années chacune, de la carrière d'un salarié). Étudions la manière dont le consommateur choisit les quantités consommées d'un même bien composite au cours des deux périodes : la période 1 (ou période actuelle) et la période 2 (ou période future). Nous notons c_1 et c_2 les quantités consommées de cet unique bien composite. Nous notons p_1 et p_2 les prix du bien aux deux périodes. Les revenus perçus par l'individu à chaque période sont R_1 et R_2 supposés exogènes. Supposons enfin que l'individu puisse emprunter ou prêter de l'argent à un taux d'intérêt unique r. Cette hypothèse est irréaliste car jamais les banques ou établissements de crédits ne prêtent de l'argent à un taux identique à celui auquel ils rémunèrent les dépôts, même s'il s'agit de dépôts qui seront placés (et donc peu liquides). Les banques font, *a minima*, supporter le coût de leur intermédiation financière (coûts de personnel, coûts de structures, marge) et servent une rémunération aux déposants qui est toujours plus faible que la charge qui est exigée aux débiteurs. On pourrait « aisément » différencier le taux d'intérêt exigé au consommateur qui emprunte et celui qui lui est servi si, au contraire, il épargne, mais cela, sans apporter d'enseignement majeur, brouillerait les résultats principaux de l'arbitrage intertemporel de consommation. Nous notons E le montant d'épargne ($E \geq 0$) ou d'emprunt ($E \leq 0$) réalisé par l'agent en début de période 1. Cette astuce de formalisation permet de simplifier l'écriture des contraintes budgétaires de l'agent. Car il existe, non

plus une, mais deux contraintes budgétaires (une pour chaque période). Écrivons ces contraintes :

En période 1, l'exigence de ce que les dépenses n'excèdent pas les ressources s'écrit :

$$p_1 c_1 \leq R_1$$

Nous modifions cette écriture pour faire apparaître l'**épargne** ou **emprunt**, véritable courroie de transmission entre les deux périodes. L'épargne se définit comme de la « non-consommation » de 1re période, c'est-à-dire l'écart entre les ressources de 1re période et la dépense de consommation en valeur : $E = R_1 - p_1 c_1$. Remarquons que E peut être négatif, ce qui correspond bien à un emprunt auquel devrait recourir un consommateur qui dépenserait plus que le niveau de ses ressources de 1re période.

En période 2, en plus de son revenu de la période R_2, le consommateur dispose de son épargne, principal et intérêts, c'est-à-dire du montant E qui avait été placé au taux d'intérêt r et des intérêts de ce placement $r \times E$. Ainsi, il vient s'ajouter à R_2, le montant $(1 + r)E$. La contrainte budgétaire de seconde période s'écrit donc :

$$p_2 c_2 \leq R_2 + (1 + r)E$$

Pour parvenir à l'expression d'une contrainte budgétaire réunissant les exigences relatives aux deux périodes, ce que l'on désigne par contrainte budgétaire intertemporelle, nous remplaçons, dans cette seconde contrainte, E par sa valeur exprimée plus haut. Ainsi, on a :

$$p_2 c_2 \leq R_2 + (1 + r)(R_1 - p_1 c_1)$$
$$\Leftrightarrow p_1 c_1 (1 + \mathrm{r}) + p_2 c_2 \leq R_1 (1 + r) + R_2$$

Cette contrainte est la contrainte budgétaire intertemporelle exprimée en *valeur future*. On peut la réécrire sous une autre forme (en multipliant les deux membres de l'inéquation par $\frac{1}{1+r}$) :

$$p_1 c_1 + \frac{p_2 c_2}{1+r} \leq R_1 + \frac{R_2}{1+r}$$

Il s'agit de la contrainte budgétaire intertemporelle exprimée en *valeur présente* (ou actuelle) qui dispose que la somme actualisée (au taux r) des dépenses de consommation ne doit pas excéder la somme actualisée des ressources.

Il est possible de représenter cette contrainte de richesse dans le plan (O, c_1, c_2). Il s'agit du plan pour lequel figure, en abscisse, la consommation aujourd'hui et, en ordonnée, la consommation demain. Un point (c_1, c_2) est donc **une trajectoire de consommation** dans le temps. En réécrivant finalement la contrainte comme $c_2 \leq \frac{R_1(1+r)+R_2}{p_2} - \frac{p_1(1+r)}{p_2}c_1$, on délimite un ensemble budgétaire (intertemporel) représenté sur la figure 8 :

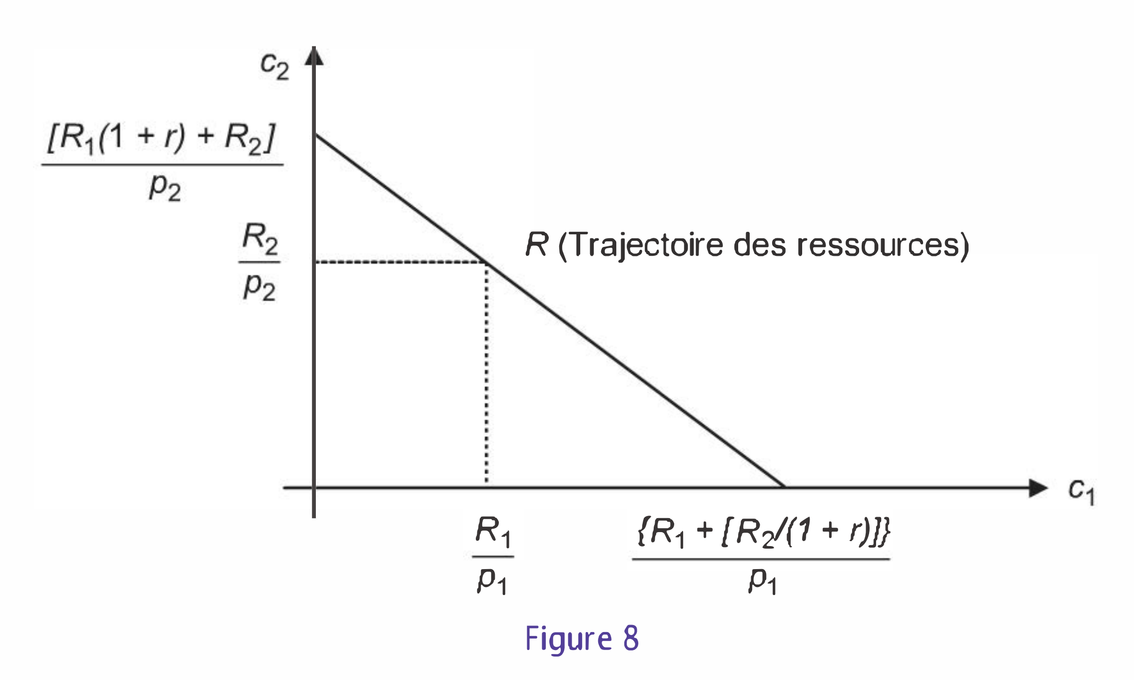


Figure 8

La droite de budget passe par le point de dotations initiales ou trajectoire des ressources R. On remarque qu'en abscisse le niveau de consommation de première période est mesuré en valeur présente, tandis qu'en ordonnée le niveau de consommation de seconde période est mesuré en valeur future. Pour établir ce que sera la trajectoire de consommation optimale, il faut maintenant confronter cet ensemble budgétaire avec les préférences temporelles du consommateur. Nous supposons que la relation de préférence ou indifférence du consommateur (définie sur des trajectoires de consommation du bien composite de la période 1 à la période 2) peut être traduite numériquement par fonction d'utilité intertemporelle notée

$u(c_1, c_2)$. On fait donc apparaître un faisceau de courbes d'indifférence qui représente les préférences du consommateur dans le repère $(0 ; c_1 ; c_2)$. Comme précédemment, la décision optimale du consommateur (la trajectoire optimale de consommation) correspond à la plus haute courbe d'indifférence compatible avec l'ensemble budgétaire intertemporel caractérisé ci-dessus. Sauf cas fortuit, la décision optimale du consommateur ne correspond pas au point *R* qui incarne la trajectoire effective des ressources. Avec des préférences strictement convexes, la décision optimale sera une solution intérieure. Deux configurations peuvent se présenter (figure 9).

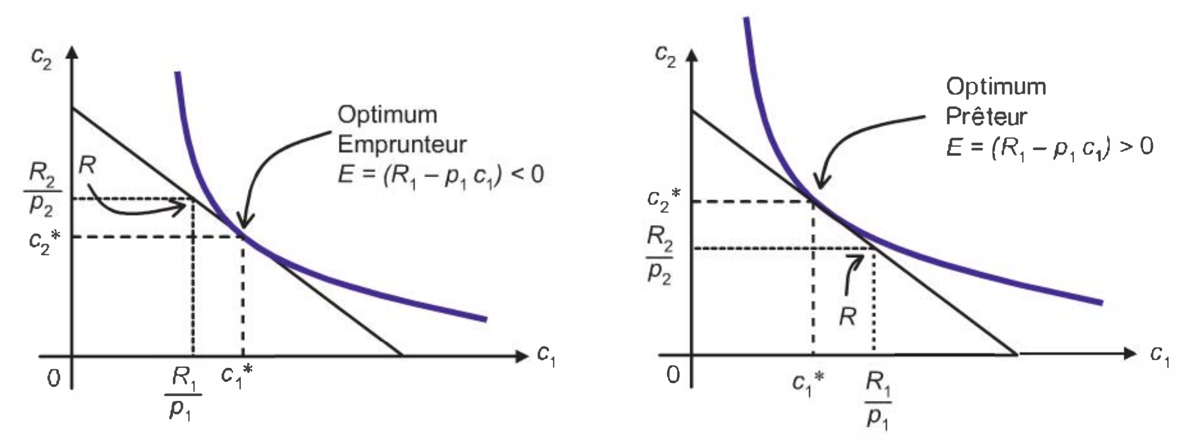


Figure 9 Les deux configurations d'optimum

Dans le premier cas, la décision optimale conduit le consommateur à consommer plus, en 1re période, que ses ressources courantes (et donc moins, en période 2, que ses ressources courantes). Il doit donc emprunter pour atteindre une trajectoire de consommation qui maximise sa satisfaction.

Dans le second cas, le consommateur consomme moins, en période 1, que ses ressources courantes. Il dégage un certain montant d'épargne qu'il va convertir en supplément de dépenses de consommation en période 2.

On remarque qu'avec des préférences strictement convexes, la décision optimale consiste à choisir une trajectoire de consommation plus régulière, plus lisse, que celles des ressources. Par exemple, si le revenu est beaucoup plus fort en période 2 qu'en période 1, il est optimal

d'emprunter beaucoup d'argent pour être à même d'avoir un train de vie, dès la période 1, assez proche de celui de la période 2. On peut illustrer ce phénomène en représentant la chronique des ressources et la chronique des dépenses de consommation dans le temps. Pour plus de réalisme, nous supposons que le consommateur décide de sa trajectoire optimale de consommation, non pas sur deux, mais sur T périodes. Avec des préférences strictement convexes, cette trajectoire pourrait ressembler à ceci (figure 10).

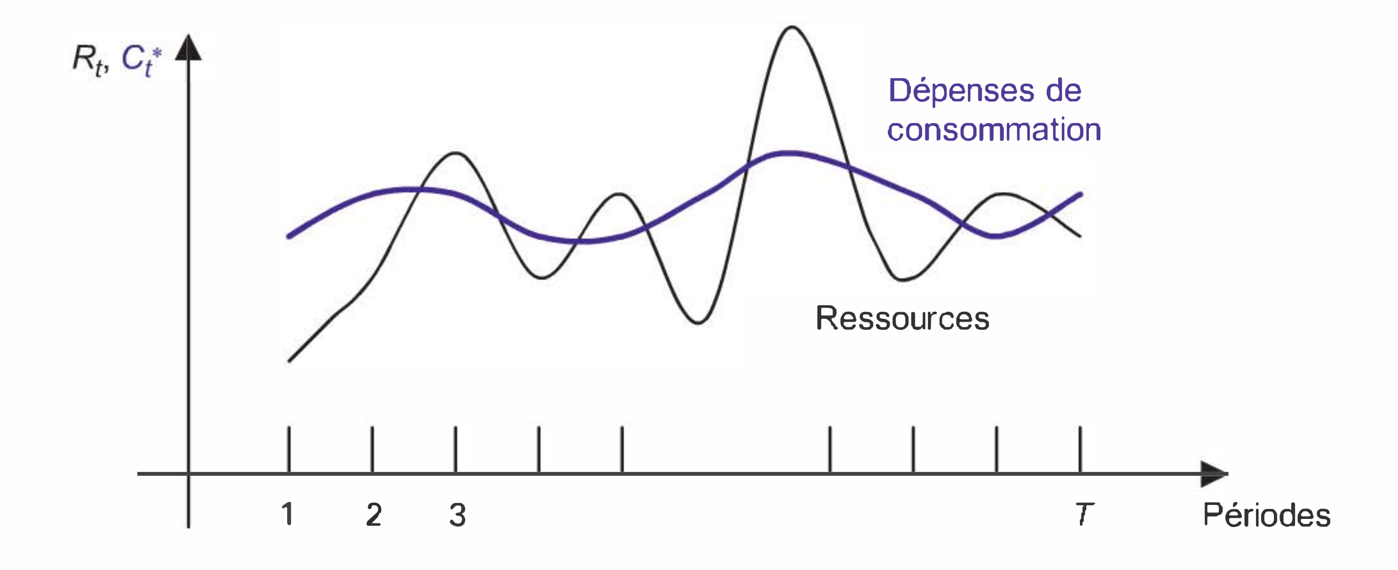


Figure 10

L'enseignement que l'on peut tirer de cette analyse est donc bien que la convexité des préférences est la marque d'un goût pour la régularité, dans le temps, du « train de vie ».

À l'inverse, des préférences strictement concaves sont la marque d'une préférence extrême pour une période particulière de la vie. En schématisant, de pures « cigales » et de pures « fourmis » ont des préférences strictement concaves : la pure cigale préfère l'été (la période 1) tandis que la pure fourmi accumulera tout l'été pour profiter de l'automne (la période 2). Dans le cas où les préférences du consommateur ne sont pas convexes, la condition équi-marginale ne sera, bien entendu, pas nécessairement vérifiée à l'optimum. Prenons précisément le cas de préférences strictement concaves et supposons que la consommation d'une période puisse

être « non nécessaire » (par exemple $c_2 \geq 0$). Il est alors envisageable de parvenir à une solution en coin telle que $c_1^* = \dfrac{R_1 + \dfrac{R_2}{(1+r)}}{p_1}$ et $c_2^* = 0$ (figure 11).

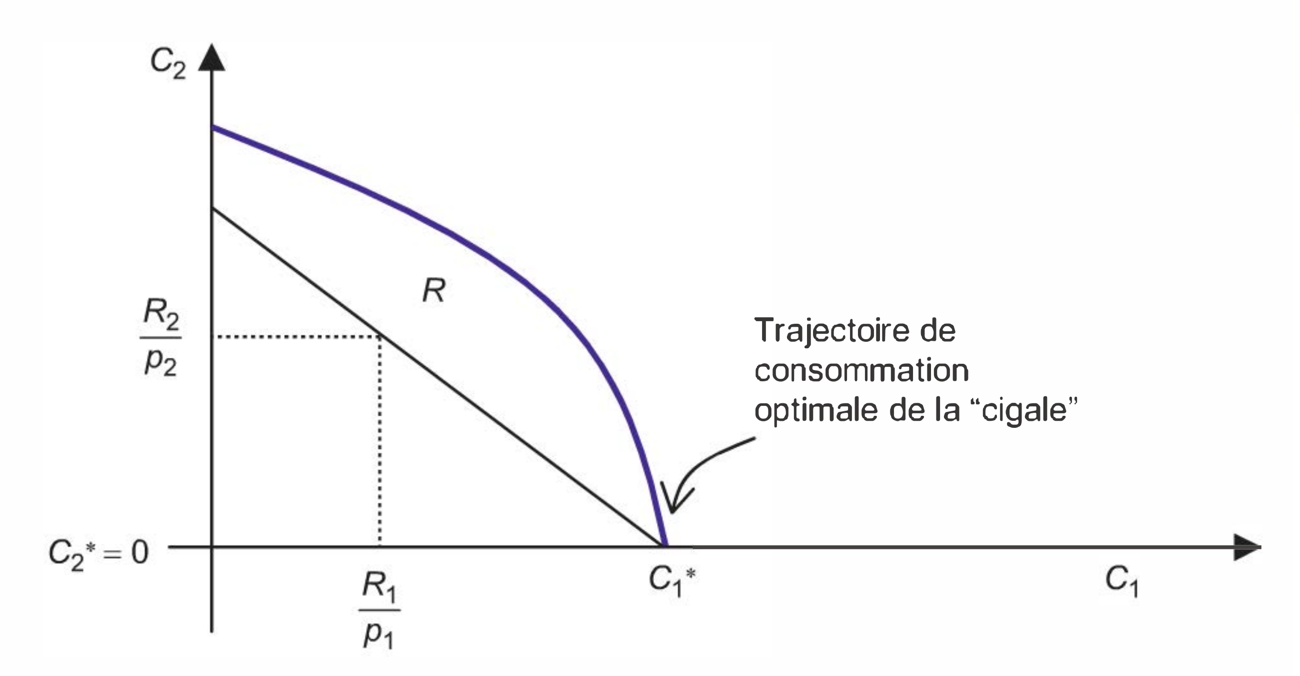


Figure 11

Un tel comportement serait celui d'un « flambeur », qui dépenserait, dès la première période, toutes les ressources obtenues durant sa vie. Cette cigale n'aurait, en conséquence, plus rien à se mettre sous la dent, une fois la bise venue. Un tel comportement est exceptionnel. Comme nous l'avons mentionné dans le chapitre précédent, il peut être le fait d'un consommateur victime d'une addiction (à la drogue, aux jeux de casino, etc.). Remarquons que prêter à un tel consommateur est très risqué car celui-ci aura peu d'incitation à réellement rembourser en période 2 : il aurait, en vérité, intérêt à « disparaître » à la fin de la période 1. C'est pourquoi les banques cherchent à détecter les candidats aux crédits qui seraient sujets à des comportements addictifs. Elles savent que ce type de débiteurs est mauvais payeur.

L'ensemble des développements de cette section repose sur l'hypothèse d'absence d'incertitude : ainsi le « flambeur » dont nous observons le

comportement ci-dessus sait précisément qu'il obtiendra le niveau de ressources R_2 en période 2 et il emprunte jusqu'à la dernière extrémité du montant envisageable. Cette hypothèse nous fait toucher les limites de l'analyse de l'arbitrage intertemporel de consommation : dans la réalité, les consommateurs ont plutôt une inclination à épargner, ne serait-ce que par précaution, c'est-à-dire en raison de la présence d'incertitude. C'est précisément cette dimension incertaine que nous allons aborder dans la prochaine section.

Pour aller plus loin

Résolution du programme du consommateur

On peut caractériser formellement la décision optimale du consommateur. Celui-ci doit résoudre le programme suivant :

$$\underset{c_1;c_2}{Max} \quad u(c_1, c_2)$$

$$s.c. \quad p_1c_1(1+r) + p_2c_2 \leq R_1(1+r) + R_2$$

La contrainte « budgétaire » du consommateur sera saturée et sa décision optimale (c_1^*, c_2^*) sera caractérisée par $TmS_{c_1/c_2} = -\frac{\frac{\partial u}{\partial c_1}}{\frac{\partial u}{\partial c_2}} = -(1+r)\frac{p_1}{p_2}$.

Ceci peut se réécrire comme $\frac{\frac{\partial u}{\partial c_1}}{(1+r)p_1} = \frac{\frac{\partial u}{\partial c_2}}{p_2}$ (condition équi-marginale) : le surcroît de satisfaction par euro dépensé obtenu grâce à la dernière unité consommée de bien *aujourd'hui* (exprimé en valeur présente) doit être égal au surcroît de satisfaction par euro dépensé obtenu grâce à la dernière unité consommée de bien *demain* (exprimé en valeur future). En d'autres termes, le consommateur doit faire coïncider son taux d'escompte personnel avec le taux d'intérêt auquel il peut prêter ou emprunter auprès d'un établissement de crédit, inflation prise en compte.

3 La décision de consommation en présence d'incertitude

La prise en compte de l'incertitude nous conduit à développer la notion de paniers de biens **contingents**. C'est sur ce type de paniers que porte la décision du consommateur et formellement, la caractérisation de sa décision optimale n'est pas différente de ce que nous avons observé jusqu'ici.

On appelle **bien contingent** un bien consommé pour une issue particulière d'une épreuve aléatoire, ou, selon la terminologie probabiliste, pour un *état de la nature* particulier. Un panier de biens contingents est un panier qui décrit les quantités consommées d'un ou plusieurs biens dans deux ou plusieurs états de la nature.

Le terme « épreuve » doit être entendu au sens large : la vie de chaque individu est, par essence, une « épreuve » aléatoire. Ainsi, la consommation d'une crème glacée au caramel beurre salé n'est pas, pour nombre d'entre nous, assimilable à une « épreuve » (ce serait plutôt un moment de délectation ultime), mais si l'on envisage la dégustation d'un tel délice *selon qu'il pleut* ou *selon qu'il fait beau*, nous construisons un panier de biens contingents associé, par nature, à une épreuve aléatoire (l'aléa est, en l'occurrence, météorologique). Et comme chacun peut convenir que la satisfaction retirée de la consommation de la crème glacée *s'il pleut* n'est pas nécessairement la même que *s'il fait beau*, il est aisé de concevoir qu'une fonction d'utilité puisse être définie sur des paniers de biens contingents.

Pour appréhender la décision du consommateur en présence d'incertitude, il faut d'abord formaliser l'existence de différents états de la nature. La coutume veut que l'on note $s_1, s_2, \ldots, s_n$ les différents états de la nature. On peut alors envisager la consommation d'un unique bien composite (identifié par la lettre c) dans ces différents états de la nature. Dès lors, c_1, c_2, …, c_n désignent les quantités consommées de ce bien dans les différents états de la nature. Dans le cas où $n = 2$, nous pouvons facilement représenter

graphiquement tout panier de biens contingents (c_1, c_2) dans un repère où c_1 figure en abscisse et c_2 figure en ordonnée. La quantité c_1 est la quantité consommée du bien si l'état de la nature s_1 se réalise et la quantité c_2 est la quantité consommée du bien si l'état de la nature s_2 se réalise (on dit aussi, respectivement, quantité consommée si s_1 et quantité consommée si s_2). Comme nous le constatons sur la figure 12, en un panier $C = (c_1, c_2)$ passe une unique courbe d'indifférence.

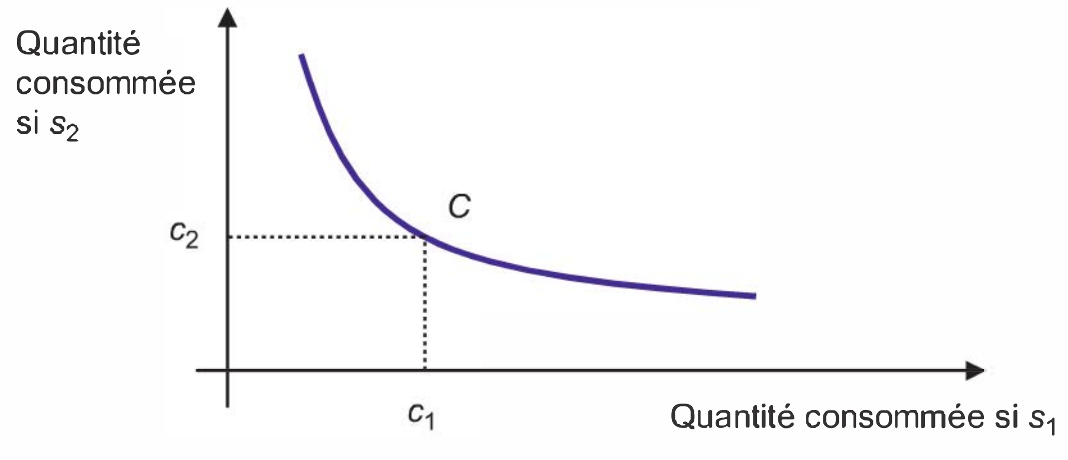


Figure 12

Faire ainsi apparaître une courbe d'indifférence dans le repère $(0 ; c_1 ; c_2)$ laisse entendre qu'il existe une fonction numérique $u(c_1, c_2)$ qui traduit convenablement les préférences du consommateur en contexte d'incertitude. Les propriétés de la fonction $u(c_1, c_2)$ dépendent des axiomes de comportement retenus et la question des « bons » axiomes à postuler n'est pas une mince affaire. Dans l'application qui suit, nous optons pour une fonction représentative des préférences assise sur un axiome appelé axiome « d'indépendance » (ou *mixture independence*) qui dispose que si un individu préfère une distribution de probabilité L_X à une autre notée L_Y, il préférera nécessairement toute combinaison linéaire convexe de L_X avec une troisième distribution de probabilité L_Z plutôt que la combinaison linéaire convexe de L_Y avec L_Z. Si un tel postulat est vérifié, la fonction d'utilité sur laquelle il convient de s'appuyer en contexte d'incertitude est dénommée **espérance d'utilité** (EU). La discussion autour des limites du modèle EU dépasse le cadre de cet ouvrage, mais il est opportun de savoir que l'axiome d'indépendance n'est pas un axiome parfaitement robuste.

L'application suivante est une décision de consommation un peu différente de ce que nous avons rencontré plus haut : il s'agit désormais de caractériser la demande d'assurance d'un consommateur. Plutôt que de raisonner sur des consommations au sens usuel c_1 et c_2, nous considérons la richesse finale du consommateur w dans deux états de la nature. L'état s_1 correspond à la réalisation d'un sinistre (incendie, accident matériel, vol, etc.) et l'état s_2 correspond à l'absence de sinistre. Nous raisonnons donc sur les variables w_1 (richesse finale en cas de sinistre) et w_2 (richesse finale en cas de non sinistre). Nous représentons le faisceau des courbes d'indifférence du décideur dans le repère (0 ; w_1 ; w_2). Ces courbes sont convexes sous une hypothèse qui peut être interprétée comme de l'aversion pour le risque dans le modèle EU. Comme précédemment, la satisfaction est d'autant plus élevée que l'on se situe vers le haut et la droite (figure 13).

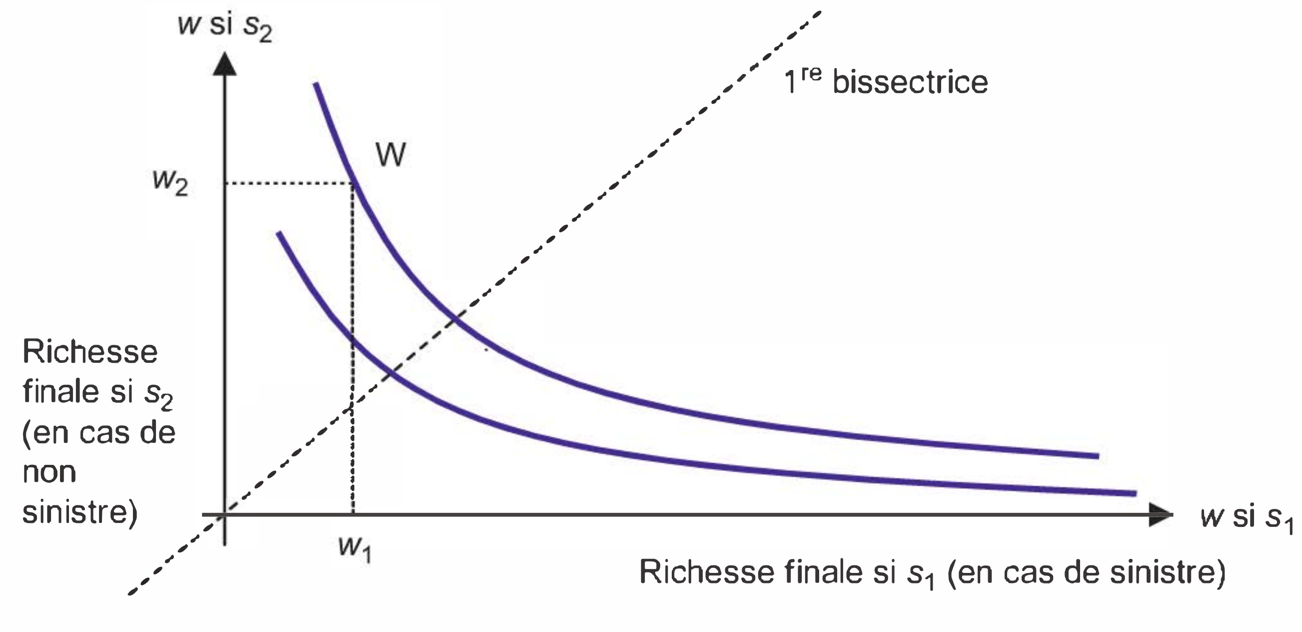


Figure 13

En l'absence d'assurance (c'est-à-dire si le consommateur ne souscrit aucun contrat), sa richesse finale est identifiée par le panier W situé au-dessus de la première bissectrice. En effet, le niveau de la richesse finale en cas de sinistre (w_1) est plus faible que le niveau de la richesse finale en cas de non sinistre (w_2) ; or, tout panier (w_1 ; w_2) pour lequel $w_2 \geq w_1$ est situé au-dessus de la première bissectrice dans le repère (0 ; w_1 ; w_2). La différence

entre w_2 et w_1 est le niveau de la perte causée par le sinistre, que l'on note généralement L (pour *loss*) : $L = w_2 - w_1$.

Le consommateur va-t-il se contenter du niveau de satisfaction atteint au point W, c'est-à-dire en l'absence de toute couverture assurantielle ? Pour le savoir, introduisons un contrat d'assurance Z caractérisé par le couple $(I\,;P)$ où I désigne le montant de l'indemnité en cas de sinistre et P le montant de la prime d'assurance.

En l'absence de sinistre, la richesse finale de l'assuré sera :

$$w_2 - P$$

En présence de sinistre, la richesse finale de l'assuré sera :

$$w_1 - P + I = w_2 - L - P + I$$

On peut faire apparaître, sur la figure 14, le panier correspondant à la richesse finale du décideur ayant souscrit au contrat d'assurance Z.

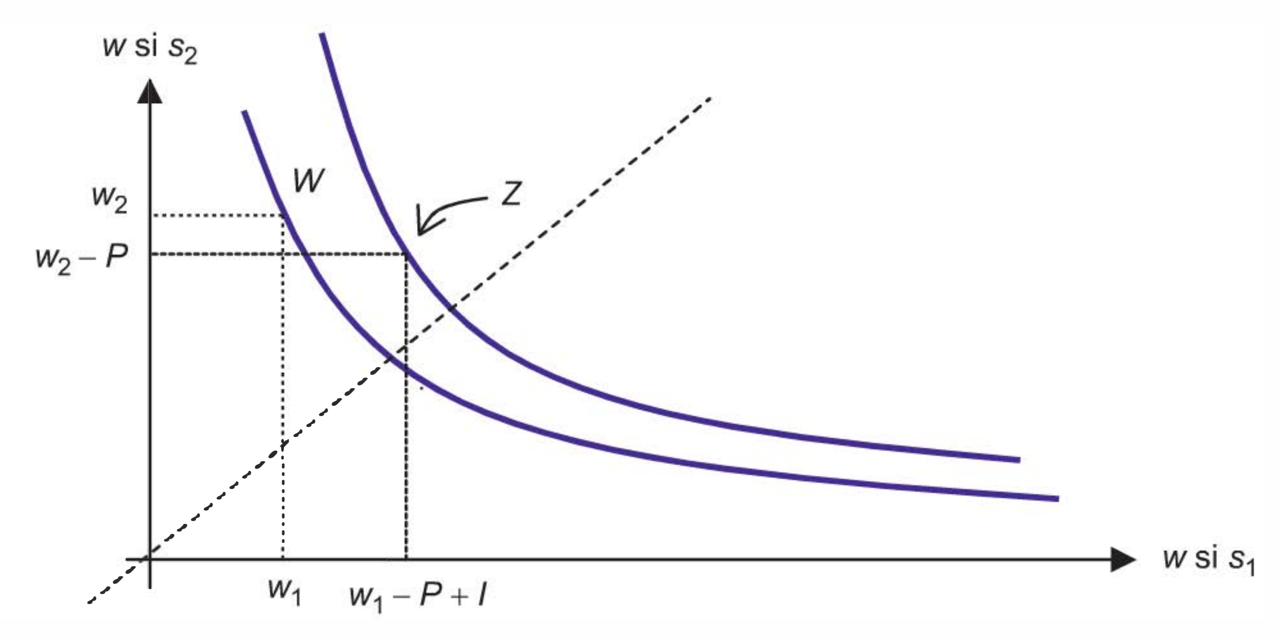


Figure 14

Le panier Z apporte un niveau de satisfaction supérieur à celle obtenue avec le panier W : le décideur choisit de souscrire au contrat d'assurance plutôt que de rester exposé au risque de sinistre sans couverture assurantielle. La décision de souscrire un contrat d'assurance dépend bien sûr du niveau de l'indemnité (en cas de sinistre) au regard de la prime. Si la prime

d'assurance est trop élevée, il se peut que l'individu, même adversaire du risque, renonce à la couverture assurantielle. Représentons, sur la figure 15, un panier Z' correspondant à un contrat d'assurance $(I\,;P')$ où la prime P' serait désormais trop onéreuse pour que le consommateur consente à souscrire au contrat.

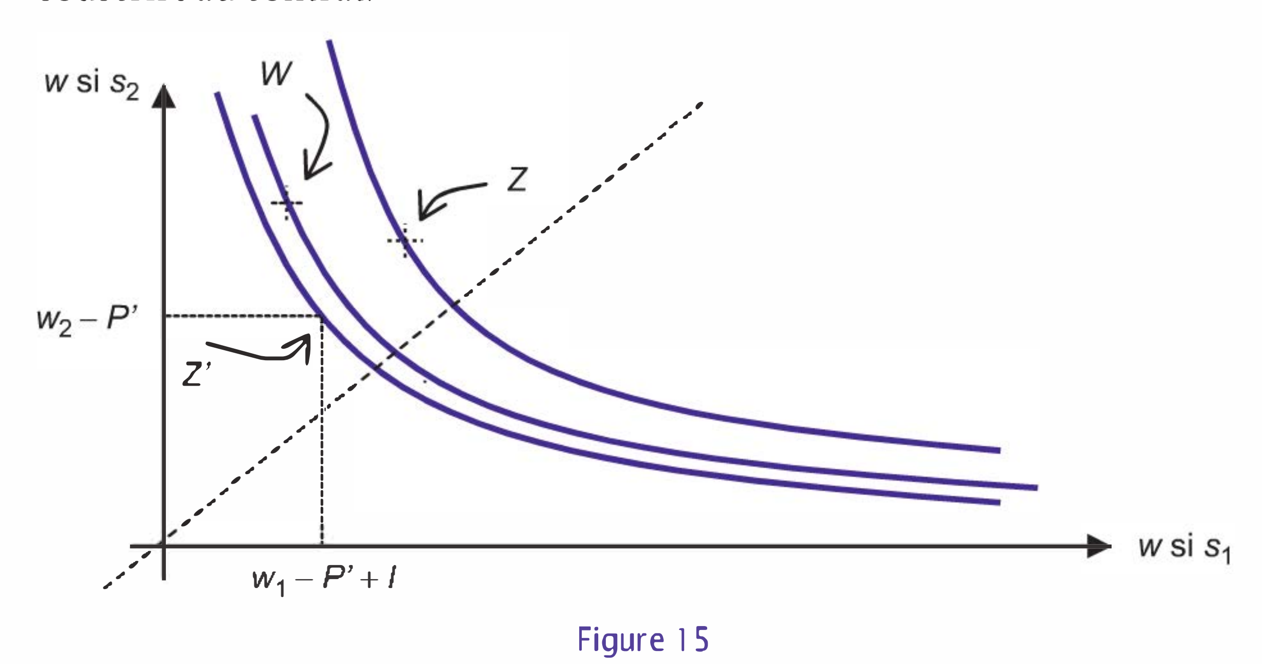


Figure 15

En Z' le niveau de satisfaction atteint par le consommateur est plus faible qu'en W. Il n'a donc pas intérêt à s'assurer.

En conclusion de ce chapitre, précisons que les différentes voies d'approfondissement évoquées dans les trois sections (consommation et loisir, intertemporel, incertitude) peuvent être mêlées. Ainsi, il est envisageable de caractériser la décision optimale du consommateur lorsqu'il constitue des paniers contenant n biens et services auxquels il faut ajouter le bien « loisir », pour m périodes distinctes et dans k états de la nature différents. Il s'agirait ainsi de constituer, de façon optimale, un panier de $(n + 1) \times m \times k$ biens. Il serait impossible de représenter graphiquement la décision optimale du consommateur, mais il est possible de résoudre mathématiquement un tel problème. Du point de vue de l'économiste, chaque décision de consommation d'un individu, à tout instant de sa vie, peut être interprété comme le résultat d'un tel effort d'optimisation, se résumant en la maximisation d'une fonction d'utilité intertemporelle et contingente (incluant le bien loisir parmi ses arguments) sous des contraintes multiples.

4 La demande globale et le surplus des consommateurs

Mots-clés

Fonction de demande, disponibilité maximale à payer, surplus des consommateurs.

Comment agréger les demandes exprimées par l'ensemble des consommateurs souhaitant consommer un bien ou service particulier ? Pour répondre à cette question, laissons de côté un des messages tiré des chapitres précédents : la demande exprimée pour un bien particulier dépend aussi du prix des autres biens. Il est entendu que la décision d'un consommateur est le résultat d'une alchimie complexe et que de multiples paramètres influencent le comportement de consommation de chacun. Ainsi, le prix d'un bien j est un élément important dans la décision de consommation d'un bien i, et ce d'autant plus si les biens i et j sont de proches substituts. Mais pour manipuler (chapitre 8) la demande en bien i globalement exprimée par l'ensemble des consommateurs, il convient, par souci de simplicité, de restreindre notre analyse aux conséquences, sur cette demande, d'une variation du seul prix du bien i. Fort heureusement, il sera toujours possible d'élargir le spectre des variables dont on jugera que la demande globale en bien i est dépendante (en particulier pour analyser un problème d'élasticité-prix croisée à l'échelle globale).

L'une des vertus de cette simplification – la demande globale en bien i n'est supposée dépendre que du prix du bien i – est de permettre, en

développant la notion de surplus des consommateurs, de construire une mesure du « bien-être » éprouvé par les consommateurs qui consomment le bien considéré.

1 La demande globale

Penchons-nous d'abord sur l'agrégation des demandes individuelles permettant de définir la demande globale d'un bien ou service. Dans le chapitre 2, nous avons établi que la demande optimale en bien *i* exprimée par un consommateur k ($k = 1, \dots, m$) est une fonction du prix du bien *i*, des prix des $(n - 1)$ autres biens et du revenu R^k du consommateur k : $x_i^k(p_1, p_2, \dots, p_i, \dots, p_n, R^k)$. Si l'on recense l'expression des fonctions de demande de bien *i* pour l'ensemble des m consommateurs, il est possible de construire la demande globale de bien comme la somme simple des demandes individuelles de ce bien.

> **La demande globale** d'un bien *i* est la somme des demandes exprimées par l'ensemble des agents participant – ou susceptibles de participer – à la consommation de ce bien. S'il y a m consommateurs acquérant ou susceptibles d'acquérir le bien, la demande globale x_i^D sera donc :
>
> $$x_i^D(p_1, p_2, \dots, p_i, \dots, p_n, R^1, R^2, \dots, R^m) = \sum_{k=1}^{m} x_i^k(p_1, p_2, \dots, p_i, \dots, p_n, R^k)$$

Intrinsèquement, la demande globale de bien *i* est une fonction de $(n + m)$ variables. Ceci n'est, à l'évidence, pas très commode à manipuler. Dans la plupart des applications qui requièrent l'utilisation explicite de cette fonction, seule la variable p_i (le prix du bien *i* lui-même) sera sujette à des variations. C'est pourquoi nous allons, dans la suite, restreindre la fonction de demande globale de bien *i* à une fonction $x_i^D(p_i)$. Cette fonction – d'une seule variable – sera, dans le cas standard, une fonction décroissante. C'est le cas correspondant à un bien **ordinaire**, déjà décrit dans le chapitre 2, et ici représenté sur la figure 1.

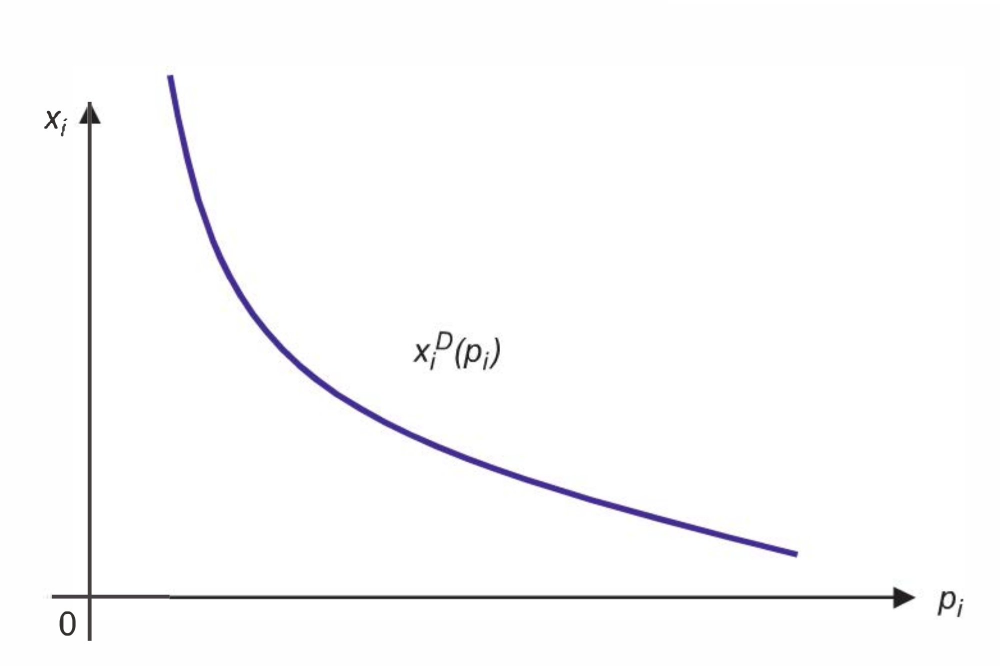


Figure 1 Demande globale d'un bien *i* ordinaire

La décroissance de cette fonction traduit le fait que la quantité globalement demandée de bien *i* par les *m* consommateurs est de plus en plus faible lorsque le prix de ce bien augmente :

- soit parce que chacun des consommateurs prenant part à la consommation du bien réduit la quantité qu'il consomme de ce bien ;
- soit parce que certains des consommateurs qui participaient initialement à la consommation du bien n'y participent désormais plus ;
- soit parce que les deux phénomènes ci-dessus décrits se produisent conjointement.

Puisque nous nous focalisons désormais sur la relation entre prix et quantité demandée d'un seul et unique bien, nous pouvons aménager la formalisation retenue. Nous choisissons de noter Q^D (plutôt que x^D) la quantité demandée du bien car la variable *Q* évoque très ouvertement une quantité et il sera plus facile d'identifier cette variable dans les chapitres où apparaîtront ensuite d'autres grandeurs.

Ce qui différencie les demandes globales des différents biens est le niveau plus ou moins élevé de leur élasticité-prix. Nous avons vu dans le chapitre 2 que l'élasticité-prix (directe) de la demande est une mesure de sensibilité de la demande au prix du bien. Dans le cas d'un bien ordinaire, l'élasticité est négative (la demande est une fonction décroissante du prix) ;

pour plus de clarté, il est de coutume de raisonner sur les valeurs absolues de l'élasticité-prix de la demande. Ainsi, lorsqu'on dit que la demande globale de bien *i* est plus élastique que la demande globale de bien *j*, cela signifie que la valeur absolue de l'élasticité-prix (directe) de la demande globale de bien *i* est supérieure à la valeur absolue de l'élasticité-prix (directe) de la demande globale de bien *j*, ce qui correspond à : $|\varepsilon_i| > |\varepsilon_j|$. L'élasticité-prix de la demande d'un bien est immédiatement visualisée en observant la pente de la courbe de demande globale. Ainsi, une demande globale peu pentue (figure 2) est le signe d'une faible élasticité-prix de la demande et une demande globale très pentue indique une forte élasticité-prix de la demande (figure 3).

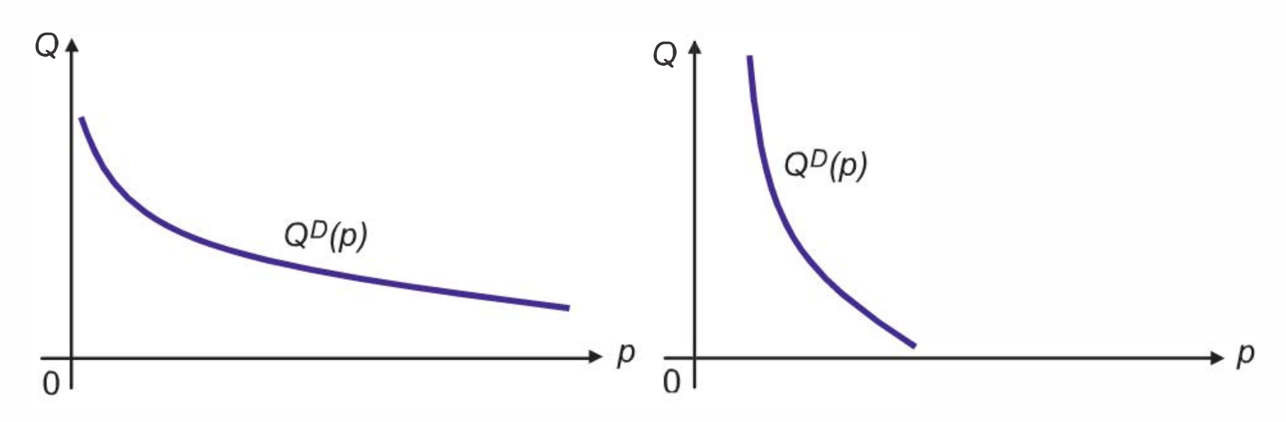


Figure 2 Demande globale faiblement élastique au prix

Figure 3 Demande globale fortement élastique au prix

Lorsque la demande globale est faiblement élastique (on dit « inélastique » au prix), une hausse substantielle du prix du bien ne fait que très peu diminuer la quantité globalement demandée. C'est le cas des carburants ou du gaz naturel consommés pour alimenter les moteurs des automobiles ou pour chauffer les logements : même si le prix du litre ou du m^3 augmente fortement, la quantité consommée ne diminue que faiblement car les consommateurs sont « prisonniers » du besoin de consommer de ces biens (pour aller à leur travail ou pour ne pas « mourir » de froid en plein hiver). Si les prix deviennent dissuasifs, la consommation finira par baisser (par exemple, les consommateurs choisiront le covoiturage ou privilégieront les transports en commun).

À l'inverse, lorsque la demande globale est fortement élastique, une petite variation à la hausse du prix du bien fait diminuer substantiellement la quantité globalement demandée : c'est le cas de certains biens alimentaires non vitaux dont le prix peu varier en fonction des saisons ou des aléas climatiques (bananes, cacao, légumes de saison, etc.).

Une confusion sur la lecture de l'élasticité-prix de la demande peut naître lorsque, au lieu d'examiner la fonction de demande (globale) directe, on se penche sur la fonction de **demande inverse**. La fonction de demande inverse est un concept similaire à celui de la fonction de demande directe au sens où il exprime la même relation fonctionnelle entre prix et quantité demandée. Mais, dans cette relation fonctionnelle, l'*image* est le prix et l'*antécédent* est la quantité. En d'autres termes, la fonction de demande inverse indique, pour un niveau donné de la demande, le prix auquel correspond un tel niveau de demande.

La fonction de demande inverse relative à un bien quelconque est la relation fonctionnelle qui indique, pour un niveau donné de la quantité demandée, le prix auquel est atteinte cette quantité demandée. En d'autres termes, il s'agit de la relation fonctionnelle entre prix et quantité demandée pour laquelle la quantité figure en abscisse et le prix figure en ordonnée. Si l'on note p le prix du bien et Q la quantité demandée de ce bien, la fonction de demande inverse est notée $p(Q)$.

Si les microéconomistes souhaitent disposer d'une fonction de **demande inverse**, c'est parce que cette version de la relation fonctionnelle entre prix et quantité demandée se prête à des interprétations très utiles. Il en va ainsi de la genèse de la notion de surplus du consommateur que nous allons présenter dans la section suivante. En revanche, l'utilisation de le demande inverse commande d'adopter une lecture graphique de l'élasticité-prix de la demande elle aussi inversée. Ainsi, une demande inverse très pentue correspond à une faible élasticité-prix de la demande (figure 4) et une demande inverse peu pentue correspond à une forte élasticité-prix de la demande (figure 5).

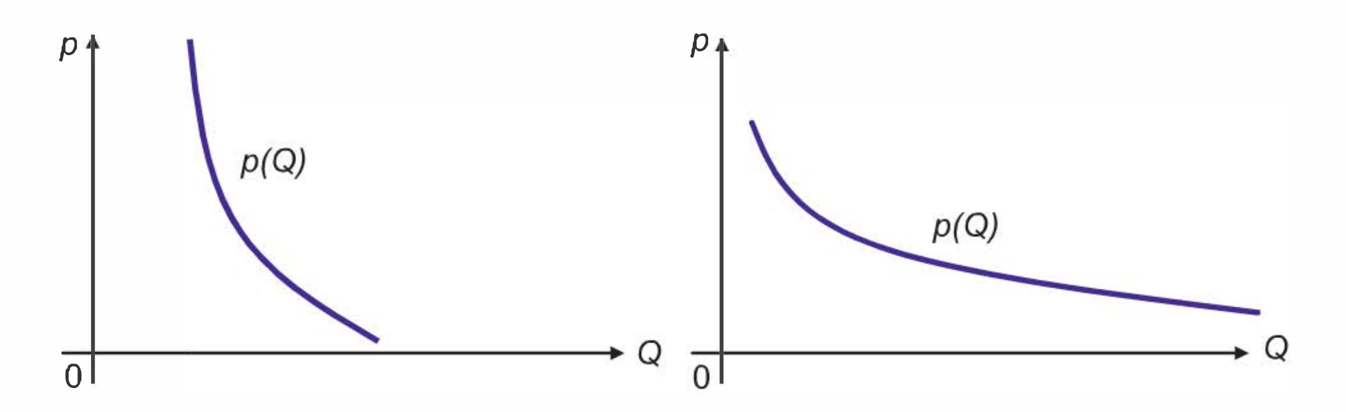


Figure 4 Demande faiblement élastique au prix

Figure 5 Demande fortement élastique au prix

2 Le surplus des consommateurs

Pour définir la notion de surplus des consommateurs, nous nous appuyons sur la fonction de demande inverse. Dans un premier temps, nous faisons l'hypothèse que le bien ou service consommé est un bien dont la consommation est unitaire, c'est-à-dire que la quantité consommée de ce bien par un individu (ou un ménage) ne peut être que 0 ou 1 : place de cinéma, trajet en avion, bicyclette, téléviseur… Dans le cas de biens dont la consommation est unitaire, la courbe de demande inverse se prête aisément à une lecture très particulière : tout se passe comme si chaque point de la courbe matérialisait la **disponibilité maximale à payer** (pour le bien considéré) d'un consommateur particulier. Nécessairement alors, les consommateurs ainsi figurés sont implicitement rangés par ordre décroissant de leur disponibilité maximale à payer. Pour s'en convaincre, examinons d'abord par une relation de demande inverse ne concernant qu'un petit nombre ($m = 3$) de consommateurs (figure 6).

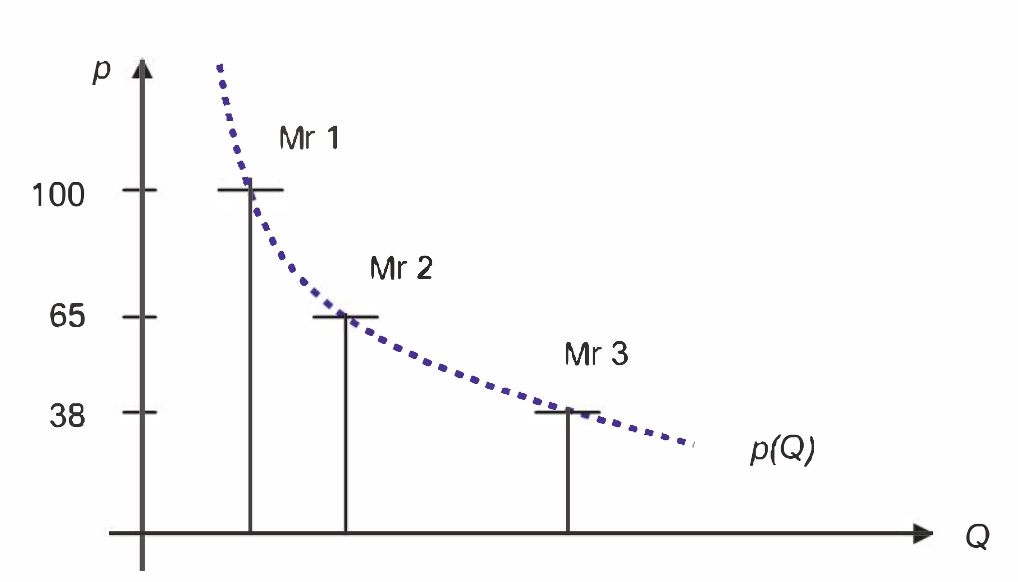


Figure 6 Consommateurs rangés par ordre décroissant de leur disponibilité maximale à payer

En toute rigueur, nous ne sommes pas ici confrontés à une fonction de demande inverse, mais à une collection de $m = 3$ points qui ébauche une courbe de demande inverse. Supposons que le bien ici consommé soit une place pour l'opéra « Les Noces de Figaro ». Le consommateur n° 1 a une disponibilité maximale à payer de 100 euros. Cela signifie qu'il est prêt à débourser une somme allant jusqu'à 100 euros pour assister au spectacle. Si le prix effectif du billet n'est que de 50 euros, il en sera ravi et ira, bien sûr, assister à l'opéra de Mozart. Le consommateur n° 3, lui, a une disponibilité maximale à payer qui n'est que de 38 euros, soit parce qu'il aime moins l'opéra, soit parce qu'il est moins fortuné, soit pour ces deux raisons conjointement. En conséquence, si le prix effectif de la place est de 50 euros, il n'achètera pas de billet et n'assistera pas à la représentation.

En pointillé, nous voyons s'ébaucher la courbe de demande inverse. Si l'on raisonne maintenant sur un groupe de $m = 5\,000$ consommateurs tous classés par ordre décroissant de leur disponibilité maximale à payer, les points ainsi figurés ne sont plus clairement distinguables les uns des autres : ils forment une courbe continue (ou presque). C'est ce que l'on voit apparaître sur la figure 7.

On observe ici que 3 500 consommateurs ont une disponibilité maximale à payer supérieure ou égale à 50 euros. Cela signifie que si le prix (supposé unique par souci de simplicité) du billet est fixé à 50 euros, 3 500 spectateurs

assisteront à la représentation d'opéra (sous réserve que la capacité d'accueil de la salle d'opéra soit supérieure ou égale à 3 500 places). La quantité apparaissant en abscisse est précisément le nombre de spectateurs présents puisque chaque consommateur n'achète qu'1 place pour l'opéra (il ne servirait à rien à un consommateur d'acheter 2 ou 3 places pour une unique représentation, quand bien même son amour de l'opéra serait immense).

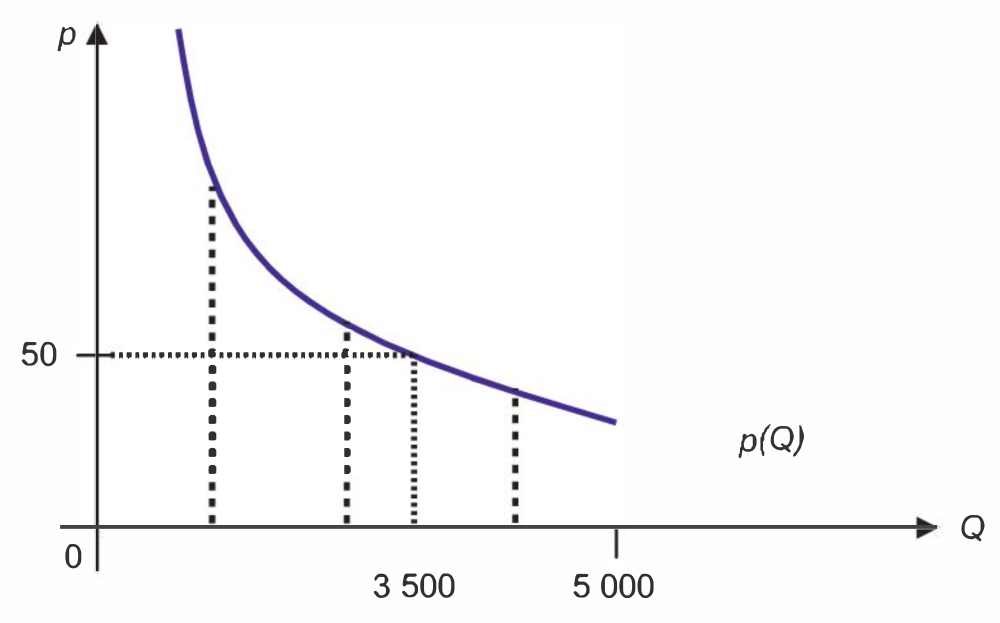


Figure 7 Continuum de consommateurs

En matière de bien-être éprouvé par les deux types de consommateurs (ceux assistant à la représentation et ceux n'y assistant pas), que pouvons-nous affirmer ? Reprenons le cas du consommateur n° 1, qui avait une disponibilité maximale à payer de 100 euros : il ne débourse que 50 euros pour assister au spectacle. Ce faisant, il réalise une sorte de gains en bien-être de 100 – 50 = 50 euros. Ce gain est certes non monétaire, mais on peut le quantifier en termes monétaires. Pour le consommateur n° 2, qui avait une disponibilité maximale à payer de 65 euros, le gain en bien-être est de 65 – 50 = 15 euros, et ainsi de suite pour tous ceux dont la disponibilité maximale à payer supérieure ou égale à 50 euros. Le consommateur « limite », c'est-à-dire celui qui a une disponibilité maximale à payer précisément égale à 50 euros, ne réalise aucun gain en bien-être : il assiste à l'opéra mais la désutilité associée au fait qu'il lui a fallu dépenser 50 euros (se privant ainsi d'autres opportunités de consommation) compense exactement la

satisfaction éprouvée en assistant au spectacle. Enfin, pour un consommateur tel que le consommateur n° 3, qui a une disponibilité maximale à payer inférieure à 50 euros, il n'y a ni gain ni perte en bien-être : il n'assiste pas à la représentation mais ne débourse aucune somme d'argent (il n'y a pas de désutilité associée à un quelconque regret ou à une quelconque frustration de ne pas assister au spectacle, en conformité avec les hypothèses postulées sur la relation de préférence ou indifférence). Ainsi, on peut définir le surplus de chacun des consommateurs en mesurant l'écart entre sa disponibilité maximale à payer et le prix effectivement payé.

Le surplus d'un consommateur quelconque tiré de la consommation d'un bien ou service particulier est défini comme l'écart entre sa disponibilité maximale à payer pour ce bien ou service et le prix qu'il paye effectivement pour l'acquérir. Dans le cas où le prix de vente du bien ou service est supérieur à sa disponibilité maximale à payer (encore appelée prix de réservation), celui-ci ne participe pas à la consommation du bien et son surplus est nul.

Le surplus d'un consommateur est un équivalent monétaire du bien-être qu'il éprouve en consommant un bien ou service. Si l'on agrège les surplus de l'ensemble des consommateurs participant à la consommation du bien, on obtient ce que l'on appelle le **surplus des consommateurs**, noté SC.

On peut remarquer que l'exemple de l'opéra proposé ci-dessus est simplificateur à l'extrême. Dans une salle d'opéra, les places ne sont pas toutes vendues au même tarif. Ainsi, de très bonnes places proches de la scène ou dans des loges confortables sont proposées à un tarif plus élevé (par exemple 90 euros). D'autres, très éloignées de la scène, sont proposées à un tarif plus bas (par exemple 35 euros). On devine l'intérêt de cette discrimination tarifaire : le consommateur n° 1 va certainement choisir une place de 1er choix à 90 euros, le consommateur n° 2 va continuer à se satisfaire d'une place à 50 euros, et le consommateur n° 3, qui était initialement exclu de la consommation du bien, va désormais pouvoir assister à la représentation car il peut acheter une place à 35 euros. Le nombre total

de spectateurs sera maintenant supérieur à ce qu'il était auparavant et, en toute vraisemblance, la recette totale tirée du spectacle sera supérieure (en réalité, il faut être sûr que le « menu » de l'offre de places à différents prix soit habilement construit au regard du profil de la courbe de demande inverse qui, rappelons-le, révèle le profil des disponibilités maximales à payer des différents consommateurs).

Revenons à notre hypothèse d'un tarif unique à 50 euros. Le surplus des consommateurs est donc la somme des écarts entre les disponibilités maximales à payer et le prix effectivement payé. Graphiquement, il s'agit de la somme de segments verticaux tels que ceux représentés sur la figure 8.

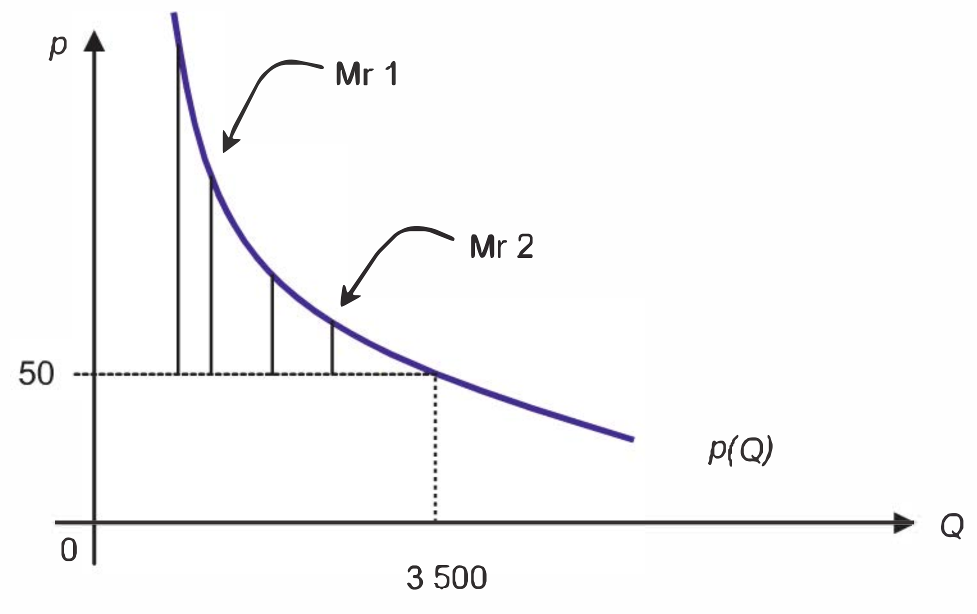


Figure 8 Surplus de quelques consommateurs

Représentons maintenant le surplus de tous les consommateurs participant à la consommation du bien (la somme des surplus individuels). Ce surplus prend la forme d'une aire, d'une surface, précisément celle située entre la courbe de demande inverse et la droite horizontale figurant un prix *p* égal à 50 euros (figure 9).

Dans le cas où la fonction de demande inverse est linéaire (c'est-à-dire prend la forme d'une droite), la surface correspondant au surplus des

consommateurs est facile à calculer : c'est l'aire d'un triangle rectangle. Par exemple, dans la figure 10, on détermine facilement que le surplus des consommateurs est égal à $\frac{p_0 - p_1}{2} \times Q_1$ (l'aire d'un triangle rectangle est la moitié de l'aire d'un rectangle).

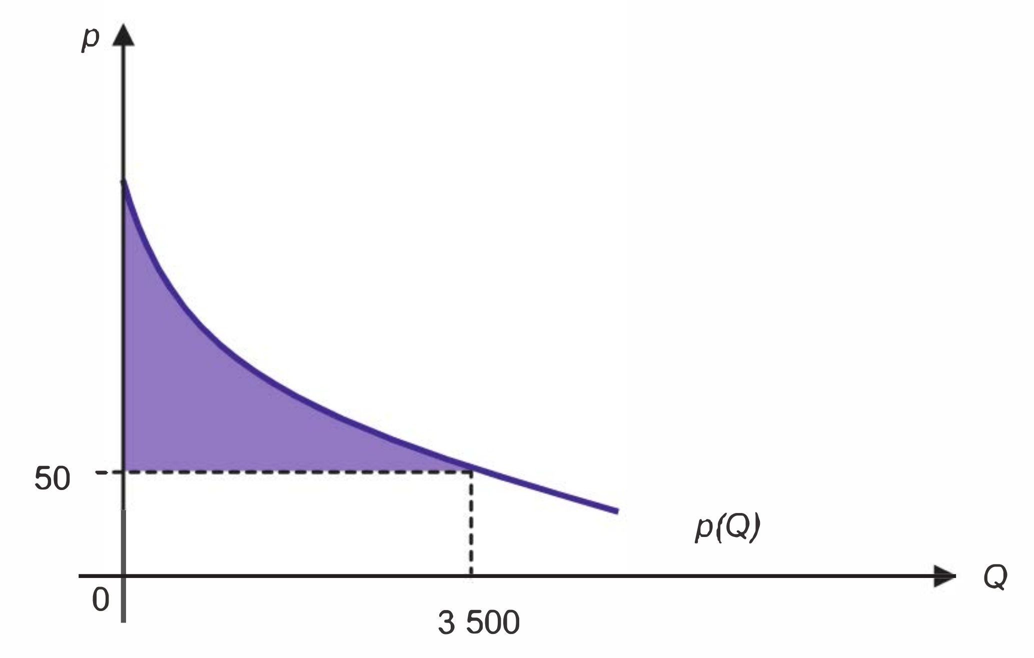


Figure 9 Surplus des consommateurs (zone mauve)

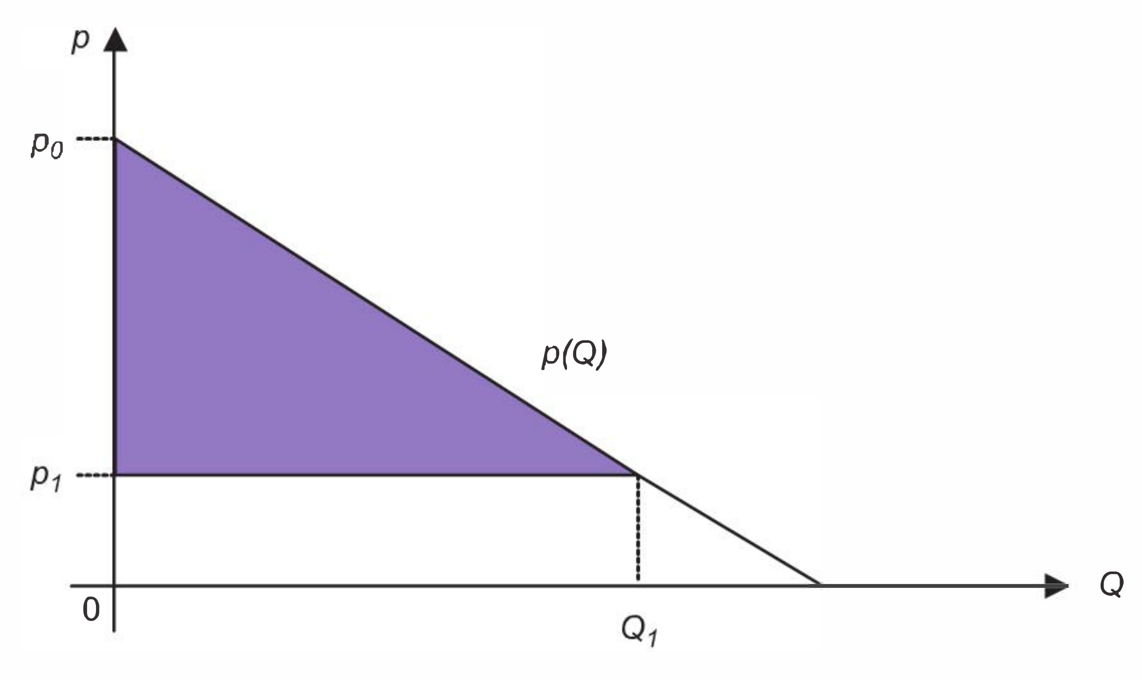


Figure 10 Surplus des consommateurs

Ce raisonnement, présenté dans le cas de biens dont la consommation est unitaire demeure valable dans le cas de biens ou services quelconques. En particulier, la surface située entre la courbe de demande inverse et la droite horizontale correspondant au prix auquel est effectivement commercialisé le bien ou service demeure égale au surplus des consommateurs.

Si nous nous référons à nouveau à la fonction de demande directe (celle pour laquelle le prix figure en abscisse et la quantité demandée figure en ordonnée), il apparaît que le surplus des consommateurs correspond à la surface sous la courbe de demande pour le seul domaine du prix où celui-ci est supérieur ou égal au prix effectivement pratiqué (figure 11).

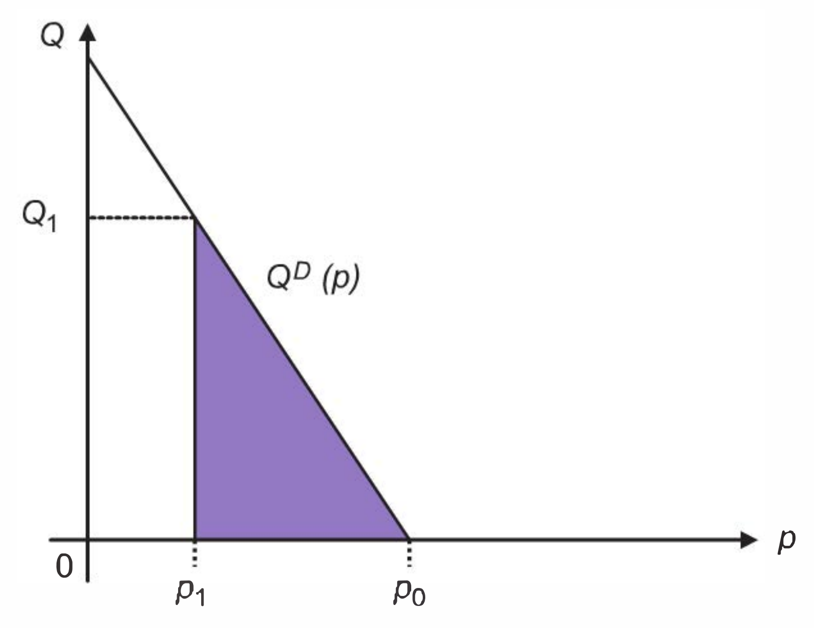


Figure 11 Surplus des consommateurs

Pour aller plus loin

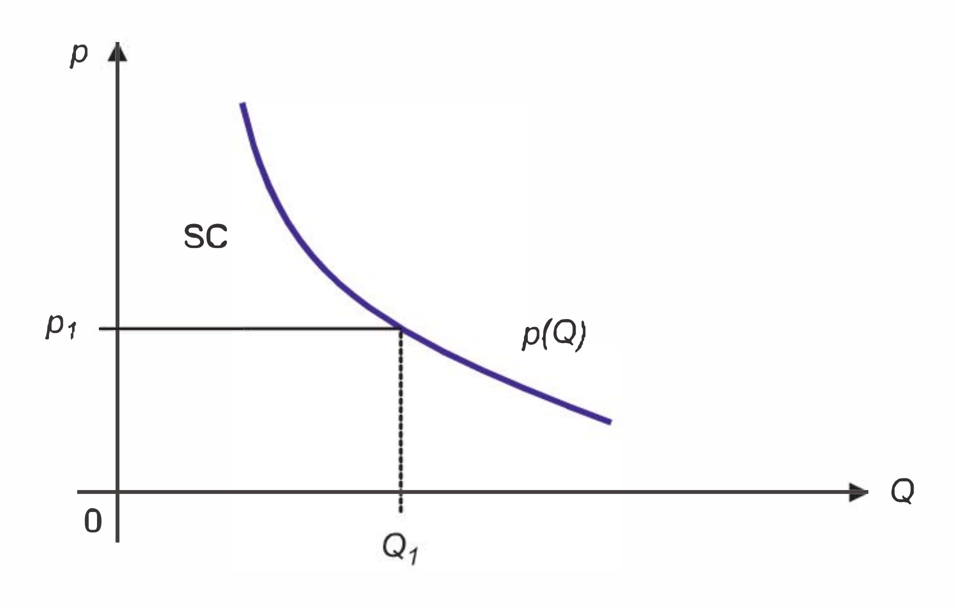
Calcul du surplus des consommateurs dans le cas d'une demande non linéaire

Raisonnons à partir de la demande inverse $p(Q)$. Supposons que le prix auquel soit commercialisé le bien ou service soit p_1 et qu'à ce prix la quantité globalement consommée soit Q_1. Le surplus des consommateurs est égal à la différence entre deux surfaces : l'aire sous la courbe de demande entre 0 et p_1 à laquelle on retranche l'aire du rectangle $(p_1 - 0) \times (Q_1 - 0) = p_1 \times Q_1$ (ce qui correspond à la dépense globalement consentie par l'ensemble des consommateurs). Sachant que l'aire sous une courbe entre deux valeurs dw'une variable est l'intégrale de la fonction entre ces deux valeurs, le surplus des consommateurs est donc (voir figure 12) :

$$SC = \int_0^{Q_1} p(Q)dQ - p_1 \times Q_1$$

Si l'on raisonne à partir de la fonction de demande directe, on peut calculer le surplus des consommateurs comme l'aire sous la courbe de demande pour le domaine du prix où celui-ci est supérieur ou égal à p_1 :

$$SC = \int_{p_1}^{+\infty} Q^D(p)dp$$



5 L'équilibre général d'une économie d'échange

Mots-clés

Boîte de Pareto-Edgeworth, optimum de Pareto, courbe des contrats, demandes excédentaires, termes de l'échange, équilibre général.

Dans l'histoire humaine, la forme primitive d'acte économique a été le troc ou échange d'un bien contre un autre (des fruits contre de la viande, des pierres taillées contre de la nourriture, du feu contre de la nourriture, etc.). Les fonctions de demande individuelles calculées dans les précédents chapitres sont elles aussi, en premier lieu, l'expression d'un désir d'échange. S'en convaincre n'est pas immédiat en raison de la présence d'un bien numéraire, c'est-à-dire d'un bien particulier en référence auquel est exprimée la valeur de tout autre bien : ce bien numéraire est la monnaie, dont le prix est égal à 1. Ainsi, il n'apparaît pas de manière parfaitement explicite que tout désir de consommation n'est que l'expression d'un désir d'échange, l'échange entre le bien convoité et un certain nombre d'unités du bien « monnaie ». Dans ce chapitre, nous revenons à cette dimension initiale de la demande de biens, et détaillons l'émergence des termes de l'échange entre deux individus désireux de procéder à l'échange d'un bien contre un autre, afin de comprendre la vraie nature du prix d'un bien ou service.

1 La boîte de Pareto-Edgeworth

Pout sonder les motivations premières, viscérales, de l'*homo œconomicus* (c'est-à-dire la femme ou l'homme rationnel), revenons à l'expression des préférences individuelles et étudions si deux individus sont susceptibles de

prendre part à un échange, en confrontant ces préférences. Ces individus sont amenés à exprimer leurs désirs de troc dans un monde où la monnaie n'existe pas et où donc il n'est pas question de revenu ou de prix au sens usuel du terme. Néanmoins, aucun des individus n'est initialement démuni. Chacun détient certaines quantités de différents biens, tels des hommes primitifs détenant des pierres taillées et de la nourriture et susceptibles d'échanger des unes contre de l'autre selon l'ampleur de leurs dotations initiales et de leurs besoins. Ces dotations initiales sont, en quelque sorte, les « ressources » de chacun : en se privant d'une partie des biens dont il est doté, l'individu s'offre l'opportunité d'obtenir en contrepartie, à l'issue de l'échange, une certaine quantité d'un ou plusieurs autres biens. L'enjeu principal, dans ce processus, est d'établir les « termes de l'échange », c'est-à-dire les proportions dans lesquelles les individus vont consentir à échanger un bien contre un autre.

1.1 Préférences et dotations

Raisonnons sur un exemple mettant aux prises deux enfants, caractérisés par des préférences différentes, possédant chacun des autos miniatures et des petits soldats et susceptibles d'échanger des uns contre des autres. Le premier enfant est identifié comme l'enfant A (parce qu'il s'appelle Anatole) et le second comme l'enfant B (car il se prénomme Brice). Les petites autos sont identifiées comme le bien indicé 1, c'est-à-dire le bien dont la quantité demandée est notée x_1 ; de même, la quantité demandée de petits soldats est notée x_2. Les préférences des deux enfants sur les paniers de ces deux biens sont incarnées par les fonctions $u^A(x_1^A ; x_2^A)$, où x_1^A et x_2^A désignent respectivement les quantités de bien 1 et de bien 2 consommées par Anatole, et $u^B(x_1^B; x_2^B)$, où x_1^B et x_2^B désignent respectivement les quantités de bien 1 et de bien 2 consommées par Brice. Connaissant ces préférences, on peut tracer les courbes d'indifférence de chacun des deux enfants, par exemple celles d'Anatole (figure 1).

Les deux enfants sont supposés posséder initialement une certaine quantité de chacun des deux biens. C'est ce que l'on appelle les dotations initiales, que nous notons « d ». Ainsi l'enfant Anatole est doté initialement d'une quantité d_1^A de bien 1 et d'une quantité d_2^A de bien 2. Quant à l'enfant Brice, il est doté d'une quantité d_1^B de bien 1 et d'une quantité d_2^B de bien 2.

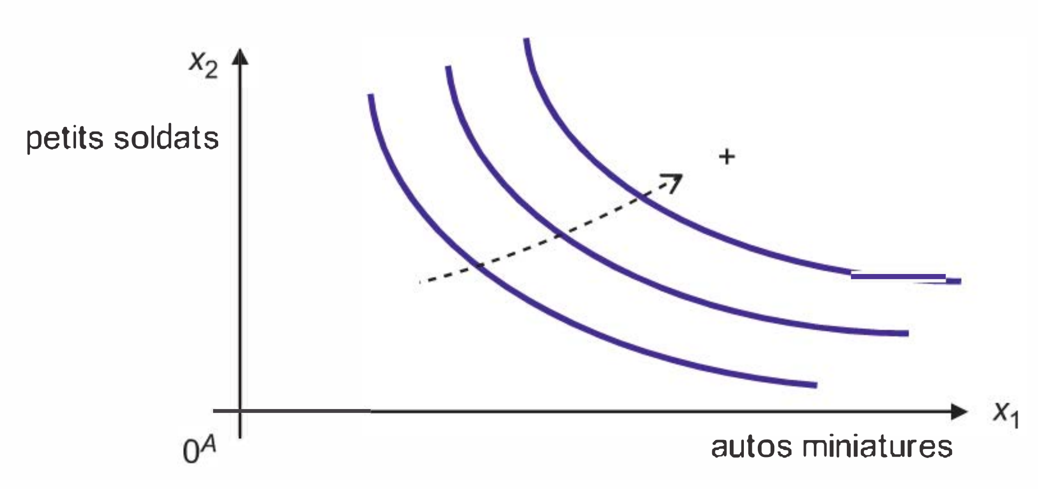


Figure 1 Préférences de l'enfant *A*

Pour faciliter la compréhension de ce modèle, construisons une application numérique simple. Supposons que :

- Anatole est doté initialement de 4 unités de bien 1 et de 3 unités de bien 2.
- Brice est doté initialement de 6 unités de bien 1 et de 2 unités de bien 2.

Ceci, dans notre formalisation, peut aussi s'écrire comme :

$$d^A = (d_1^{\,A} ; d_2^{\,A}) = (4 ; 3)$$

$$d^B = (d_1^{\,B} ; d_2^{\,B}) = (6 ; 2)$$

Commençons par représenter la dotation d'Anatole dans le repère $(0^A ; x_1^{\,A} ; x_2^{\,A})$ sur la figure 2 ; dans cette figure, nous faisons aussi apparaître la courbe d'indifférence d'Anatole passant par le point de ses dotations initiales (noté D^A) :

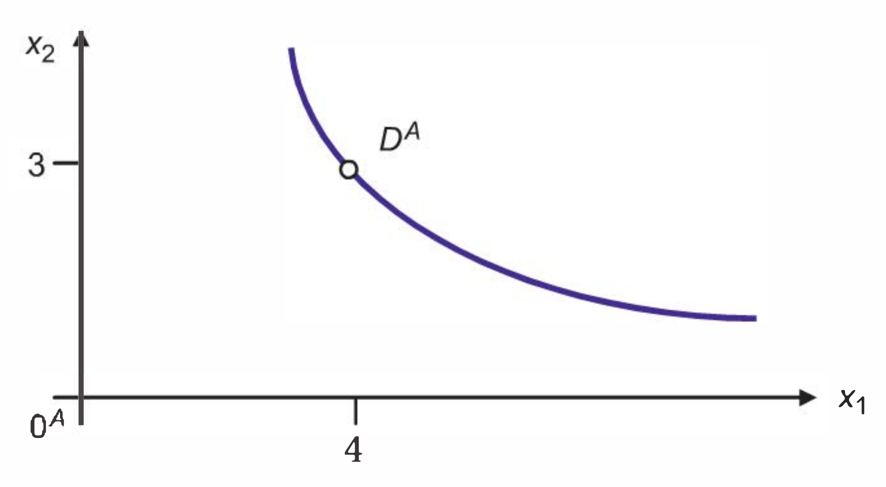


Figure 2 Dotations initiales de l'enfant *A*

Représentons ensuite la dotation de Brice dans un repère (0^B ; x_1^B ; x_2^B) sur la figure 3 ; dans cette figure, nous allons également faire apparaître la courbe d'indifférence de Brice passant par le point de ses dotations initiales (noté D^B) :

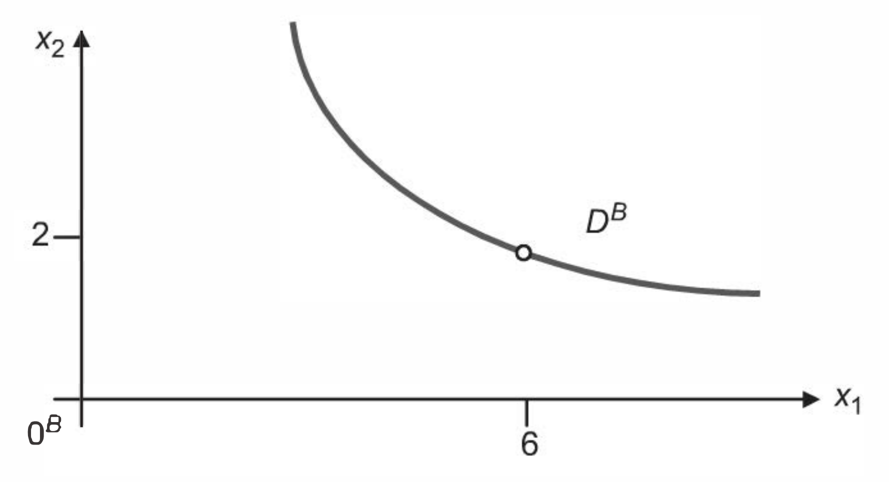


Figure 3 Dotations initiales de l'enfant *B*

1.2 Le cadre de l'échange

« L'économie » dans son ensemble est constituée de deux ressources en quantités limitées (ou rares), le bien 1, disponible globalement en quantité $d_1^A + d_1^B = 4 + 6 = 10$ unités, et le bien 2, disponible globalement en quantité $d_2^A + d_2^B = 3 + 2 = 5$ unités. Elle met en présence seulement deux agents, Anatole et Brice. C'est une économie dans son expression la plus simple. Nous allons maintenant construire la boîte de Pareto-Edgeworth représentative de cette économie. C'est une boîte rectangulaire de taille $d_1^A + d_1^B = 10$ pour le bien 1 (longueur), et de taille $d_2^A + d_2^B = 5$ pour le bien 2 (largeur). Pour la construire, on place de façon habituelle le repère orthonormé de l'enfant *A*, puis on place le repère orthonormé de l'enfant *B* « la tête en bas » (figure 4).

La boîte de Pareto-Edgeworth d'une économie (souvent appelée « boîte d'Edgeworth ») est un rectangle figurant toutes les répartitions possibles des consommations de deux biens entre deux agents économiques. La longueur et la largeur de cette boîte correspondent respectivement aux quantités totales présentes de biens 1 et 2 dans l'économie.

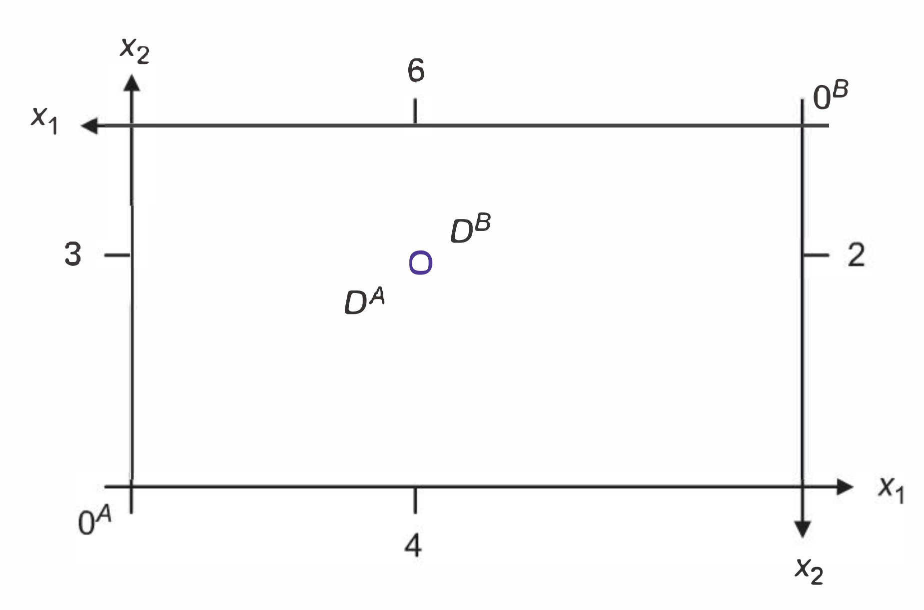


Figure 4 Boîte de Pareto-Edgeworth de l'économie

Comme on le voit très clairement sur la figure 4, le point des dotations initiales d'Anatole est, par construction, confondu avec le point des dotations initiales de Brice dans la boîte de Pareto-Edgeworth : $D^A = D^B$. Le cadre de la boîte de Pareto-Edgeworth va nous permettre de visualiser très simplement les forces qui poussent les deux enfants à échanger un bien contre un autre dans le but d'accroître simultanément leurs satisfactions respectives. Pour voir apparaître les forces en jeu, nous traçons tout d'abord les courbes d'indifférence d'Anatole et de Brice passant par le point commun de leurs dotations initiales (que nous désignerons désormais par D) sur la figure 5.

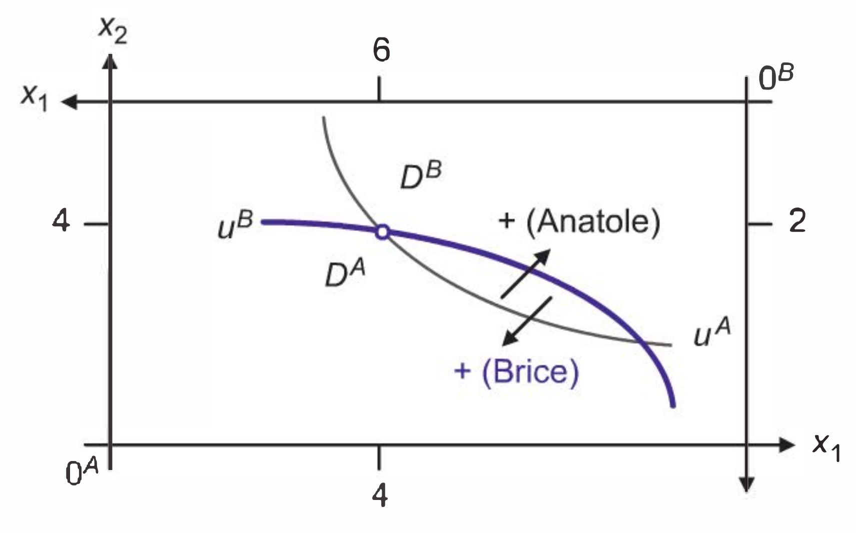


Figure 5 Courbes d'indifférences de *A* et *B* passant par le point de leurs dotations initiales

Commençons par un constat graphique : en ce point des dotations initiales D, les taux marginaux de substitution des deux enfants sont différents, ou, en d'autres termes, les proportions dans lesquelles l'un et l'autre seraient susceptibles de substituer un bien par l'autre diffèrent. Rappelons que le TmS d'un consommateur se lit immédiatement en tout point d'une courbe d'indifférence car il est égal à la pente de la tangente à la courbe d'indifférence en ce point (cf. chapitre 1) – mentionnons ici que cette propriété est également valable si l'un des agents voit ses courbes d'indifférence positionnées dans un repère « tête en bas ». Ainsi, sur la figure 6, on constate sans ambiguïté que la valeur absolue du TmS d'Anatole au point D est plus élevée que la valeur absolue de celui de Brice (au même point).

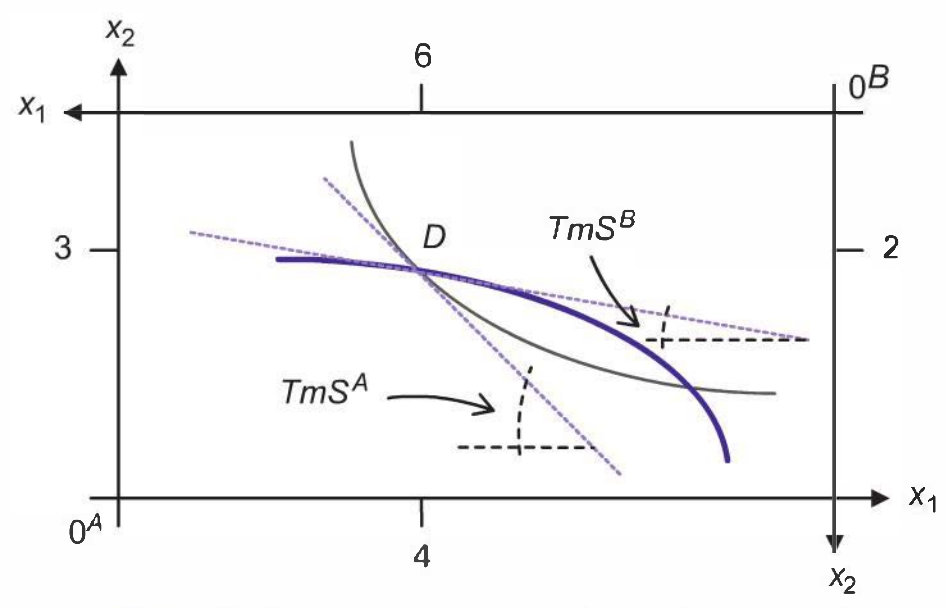


Figure 6 Taux marginaux de substitution de A et B au point de leurs dotations initiales

Une autre manière de constater graphiquement la non-coïncidence entre les TmS des deux enfants est de remarquer que les deux courbes d'indifférence forment une « lentille » dont l'un des sommets est le point D : si les TmS étaient identiques, les deux tangentes seraient confondues et aucune lentille ne se formerait. Or, en tout point de cette lentille, à la fois Anatole et Brice accroîtraient leur satisfaction au regard de la satisfaction qu'ils éprouvent s'ils consomment les quantités disponibles en D. Ceci est très clair lorsque l'on se penche à nouveau sur la figure 5 où apparaissent les sens de la satisfaction croissante pour l'un et l'autre des enfants. En d'autres termes, la non-coïncidence des TmS ou l'existence d'une lentille (qui sont

deux constats équivalents) sont le signe de l'existence d'**opportunités d'échange** entre les deux enfants. Lequel de ces échanges sera le meilleur auquel les deux enfants pourront s'adonner ?

2 La courbe des contrats

2.1 L'opportunité d'échanger

Le point *D* des dotations initiales forme le sommet d'une lentille. Cette lentille regroupe les allocations qui, si elles étaient atteintes par les consommateurs, leur permettraient d'accroître mutuellement leurs satisfactions. Pour s'en convaincre, calculons les *TmS* des enfants au point *D* et constatons qu'en ce point existe un éventail du rapport d'échange – c'est-à-dire des proportions dans lesquelles un bien peut être échangé contre l'autre – acceptable (car favorable) aux yeux des deux protagonistes. Supposons que les fonctions d'utilité des enfants A et B soient respectivement :

$$u^A(x_1^A ; x_2^A) = (x_1^A)^2 . x_2^A \text{ et } u^B(x_1^B ; x_2^B) = x_1^B . x_2^B$$

Exprimons les *TmS* d'Anatole et de Brice. Rappelons que

$TmS_{1/2} = -\dfrac{\dfrac{\partial u(x_1, x_2)}{\partial x_1}}{\dfrac{\partial u(x_1, x_2)}{\partial x_2}}$. Il est facile d'établir que $TmS_{1/2}{}^A = -\dfrac{2x_1^A x_2^A}{(x_1^A)^2} = -\dfrac{2x_2^A}{x_1^A}$

et que $TmS_{1/2}{}^B = -\dfrac{x_2^B}{x_1^B}$. Ainsi au point *D*, puisque $(d_1^A ; d_2^A) = (4 ; 3)$,

$TmS_{1/2}{}^A = -\dfrac{2 \times 3}{4} = -\dfrac{3}{2} = -\dfrac{1}{0{,}66}$ et puisque $(d_1^B ; d_2^B) = (6 ; 2)$,

$TmS_{1/2}{}^B = -\dfrac{2}{6} = -\dfrac{1}{3}$. Rappelons que de tels taux de substitution mesurent un ratio $-\dfrac{\Delta x_2}{\Delta x_1}$: au numérateur figure la variation de la quantité consommée de bien 2 et au dénominateur figure la variation de la quantité consommée de bien 1. Cela signifie qu'au point *D* des dotations initiales, Brice est prêt à céder jusqu'à 3 autos miniatures (bien 1) pour obtenir

1 petit soldat (bien 2) et que dans le même temps Anatole est prêt à céder 1 soldat s'il obtient en échange au moins 0,66 voiture. Il existe manifestement la possibilité, pour les deux enfants, de trouver un terrain d'entente : un échange est possible (Anatole va céder 1 ou plusieurs soldats à Brice en échange d'un certain nombre d'autos), dans un rapport d'échange (qui reste à déterminer) compris entre 1 (soldat) contre 3 (autos) et 1 (soldat) contre 0,66 (auto). Ce que nous venons d'établir ici est l'éventail acceptable du rapport d'échange.

Si, en revanche les *TmS* des deux enfants avaient été identiques au point *D*, aucune opportunité d'échange n'aurait existé. Ceci est un cas très particulier, qui se produit pour un certain nombre d'allocations au sein de la boîte d'Edgeworth. Ces points sont caractérisés par une propriété remarquable : à partir de telles allocations, il est impossible de procéder à un échange qui améliore la satisfaction d'un agent sans détériorer celle de l'autre. Cette propriété est la définition même d'un optimum de Pareto.

2.2 La Pareto-optimalité

Une allocation est **optimale (au sens de Pareto)** si aucun individu ne peut accroître sa satisfaction sans détériorer celle d'un autre au moins.

À l'inverse, s'il est possible de trouver une manière quelconque d'accroître la satisfaction d'un ou plusieurs individus sans pénaliser quelqu'un d'autre, il s'agit d'une allocation non optimale ; s'il n'est pas possible de trouver une telle amélioration, l'allocation est optimale.

L'optimum de Pareto est une notion d'efficacité économique et non une notion d'équité ou de justice. L'efficacité est un critère objectif au sens où n'importe quel observateur conviendra du caractère optimal d'une allocation. En revanche, il n'existe aucun critère d'équité universel : ce qui paraît équitable à un observateur pourra paraître inique à un autre. L'économiste peut néanmoins se préoccuper de questions d'équité, mais il doit au préalable définir le critère d'équité, de justice qu'il retient, en ayant conscience du caractère arbitraire de ce choix. Cette question de la multiplicité des conceptions de l'équité, aussi passionnante soit-elle, n'est pas abordée dans

ce manuel. Nous nous limitons aux questionnements sur l'efficacité paretienne (c'est-à-dire au sens de Pareto). Une telle notion est difficile à cerner et à désigner. Vilfredo Pareto suggérait même que soit utilisée une terminologie spécifique. Il utilisait le terme « maximum d'ophélémité ». Cette terminologie est malheureusement tombée en désuétude. Elle était pourtant moins équivoque que le terme d'« efficacité » aujourd'hui employé qui, s'il est relativement approprié pour qualifier la manière dont sont organisées les activités productives, n'est pas évocateur quand il s'agit de décrire les propriétés allocatives d'une économie.

Pour mieux cerner la notion d'optimum de Pareto, revenons à l'économie résumée par notre boîte de Pareto-Edgeworth et tentons d'identifier toutes les allocations qui seraient des optima de Pareto. Le point des dotations initiales *D* n'est pas un optimum de Pareto car en ce point les *TmS* des deux enfants diffèrent et qu'en conséquence existe une lentille correspondant à une zone qui, s'ils y accédaient, leur permettraient d'accroître simultanément leurs satisfactions. Si a contrario les *TmS* des deux enfants étaient identiques au point de leurs dotations initiales, il n'existerait pas une telle lentille délimitant une zone d'échanges mutuellement avantageux. Plus largement, si en une allocation quelconque au sein de la boîte d'Edgeworth, les *TmS* des deux agents sont identiques, l'absence d'opportunité de procéder à un échange mutuellement bénéfique est le signe que cette allocation est un optimum de Pareto. En d'autres termes, les optima de Pareto sont les points pour lesquels les tangentes aux courbes d'indifférence des agents sont confondues, ou, pour dire les choses plus simplement, ce sont les *points de tangence entre les courbes d'indifférence*. Représentons, sur la figure 7, l'ensemble des points de tangence entre les courbes d'indifférence.

Cette courbe (en gras) est la **courbe de tous les optima de Pareto**. Elle est encore appelée **courbe des contrats**.

Dans une économie d'échange, la **courbe des contrats** est la courbe des optima de Pareto.

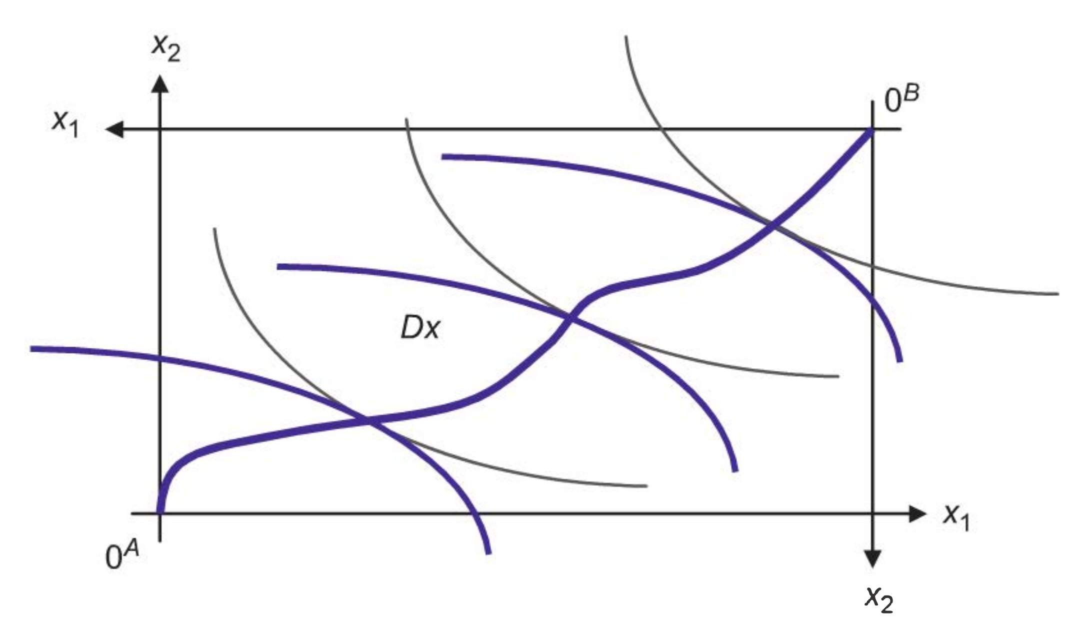


Figure 7 Ensemble des points de tangence entre les courbes d'indifférence des deux agents

Le processus d'échanges qui va se mettre en place entre les deux enfants doit logiquement conduire à une allocation située sur cette courbe des contrats. Pour déterminer quelles allocations sont optimales, nous allons devoir établir les « termes de l'échange » c'est-à-dire le rapport de l'échange entre les autos miniatures et les petits soldats auxquels les deux enfants vont spontanément procéder.

3 L'équilibre général de l'économie d'échange

Nous avons établi que le rapport d'échange devait se situer entre 1 (soldat) contre 3 (autos) et 1 (soldat) contre 0,66 (auto). Pour déterminer le rapport d'échange qui sera effectivement retenu, nous attribuons, de manière fictive, des prix p_1 et p_2 aux deux biens. Ces prix n'existent pas réellement, mais en les faisant apparaître dans notre modélisation, nous pouvons déterminer le rapport d'échange pratiqué par les enfants ; et comprendre alors ce qu'est la vraie nature d'un prix.

Déterminons tout d'abord la demande optimale de biens 1 et 2 exprimée par chaque enfant à l'issue de la maximisation de sa satisfaction sous

contrainte budgétaire. Il est, grâce à l'existence des prix p_1 et p_2, désormais possible d'écrire et de résoudre ce programme du consommateur. En particulier, nous sommes désormais en mesure d'écrire la contrainte budgétaire. Si, du côté des dépenses, elle prend sa forme habituelle, nous devons faire preuve d'astuce pour exprimer ce que sont les ressources de chaque consommateur. En réalité, dans ce modèle, les seules ressources du consommateur sont ce qu'il possède déjà sous forme de différents biens. Ce n'est qu'en cédant une certaine quantité d'un bien qu'il possède initialement qu'il peut puiser dans ses « ressources » et les convertir en la capacité à obtenir un autre bien. Ainsi donc, les ressources d'un consommateur, dans ce modèle, sont ses seules dotations initiales. Encore faut-il exprimer ces dotations en valeur (on ne peut additionner les torchons et les serviettes). Grâce aux prix ci-dessus définis, cette valorisation des ressources est possible. Ainsi les ressources d'Anatole sont la somme de ses dotations en valeur, c'est-à-dire : $p_1 d_1^A + p_2 d_2^A$. De même, les ressources de Brice sont : $p_1 d_1^B + p_2 d_2^B$. Il est donc désormais possible décrire le programme du « consommateur » A comme :

$$\begin{array}{ll} Max & u^A(x_1^A ; x_2^A) \\ s.c. & p_1 x_1^A + p_2 x_2^A \leq p_1 d_1^A + p_2 d_2^A \end{array}$$

Et un programme analogue peut être écrit pour le « consommateur » B.

En supposant que les contraintes budgétaires seront saturées (c'est-à-dire que le consommateur alloue toutes ses ressources disponibles à la consommation des biens), on peut résoudre les programmes des consommateurs A et B. Les demandes optimales de bien 1 et 2 exprimées par Anatole et Brice (en conservant les hypothèses sur les fonctions d'utilité u^A et u^B) sont :

$$x_1^A = \frac{2(p_1 d_1^A + p_2 d_2^A)}{3p_1} = \frac{8p_1 + 6p_2}{3p_1} \quad \text{et} \quad x_2^A = \frac{p_1 d_1^A + p_2 d_2^A}{3p_2} = \frac{4p_1 + 3p_2}{3p_2}$$

$$x_1^B = \frac{p_1 d_1^B + p_2 d_2^B}{2p_1} = \frac{6p_1 + 2p_2}{2p_1} \quad \text{et} \quad x_2^B = \frac{p_1 d_1^B + p_2 d_2^B}{2p_2} = \frac{6p_1 + 2p_2}{2p_2}$$

5

Ces expressions de demandes optimales de biens ont une forme un peu différente de ce que nous trouvions dans les chapitres 2 et 3. Plutôt que de nous focaliser sur ces fonctions de demandes, concentrons notre attention sur un autre outil, dérivé de celui-ci : les fonctions de **demandes excédentaires** de biens.

Dans une économie d'échange, la **demande excédentaire** d'un bien exprimée par un agent est la différence entre sa demande optimale et sa dotation initiale en ce bien.

La demande excédentaire est un outil permettant de mettre en évidence le désir d'échange du consommateur considéré en fonction du prix relatif des biens.

En exprimant les demandes excédentaires, nous mettons en effet en évidence l'intensité du désir d'échange des deux protagonistes. L'intensité de ce désir dépend bien entendu des termes de l'échange. Plus les termes de l'échange sont avantageux pour Anatole, plus celui-ci veut échanger des petits soldats contre des autos (et moins Brice le souhaitera) et inversement. Toute l'information sur les aspirations de l'un et de l'autre est présente dans l'expression des demandes excédentaires et l'élément décisif permettant de confronter quantitativement ces aspirations est le couple de prix fictifs $(p_1 ; p_2)$. En fait, nous le verrons, ce ne sont pas les prix p_1 et p_2 pris isolément qui seront déterminants, mais bien le seul ratio $\frac{p_1}{p_2}$. Calculons les demandes excédentaires en biens 1 et 2 exprimées par Anatole et Brice. Nous allons choisir de noter ces demandes excédentaires de biens à l'aide de la lettre « z ». Ainsi, par exemple, la demande excédentaire de bien 1 exprimé par Anatole (différence entre sa demande optimale et sa dotation initiale en bien 1) est $z_1^A = x_1^A - d_1^A$. On obtient donc :

$$z_1^A = x_1^A - d_1^A = \frac{8p_1 + 6p_2}{3p_1} - 4 = \frac{6p_2 - 4p_1}{3p_1}$$

$$\text{et} \quad z_2^A = x_2^A - d_2^A = \frac{4p_1 + 3p_2}{3p_2} - 3 = \frac{4p_1 - 6p_2}{3p_2}$$

$$z_1^{\,B} = x_1^{\,B} - d_1^{\,B} = \frac{6p_1 + 2p_2}{2p_1} - 6 = \frac{2p_2 - 6p_1}{2p_1}$$

$$\text{et} \quad z_2^{\,B} = x_2^{\,B} - d_2^{\,B} = \frac{6p_1 + 2p_2}{2p_2} - 2 = \frac{6p_1 - 2p_2}{2p_2}$$

Que nous apprennent ces expressions ? Pour les décrypter, il suffit de s'interroger de la manière suivante : « Dans quelles circonstances telle ou telle demande excédentaire est-elle positive, négative ? ». Commençons par la demande excédentaire en bien 1 d'Anatole. Elle est positive lorsque : $z_1^{\,A} \geq 0 \Leftrightarrow \dfrac{6p_2 - 4p_1}{3p_1} \geq 0 \Leftrightarrow 6p_2 - 4p_1 \geq 0 \Leftrightarrow 6p_2 \geq 4p_1 \Leftrightarrow \dfrac{p_1}{p_2} \leq \dfrac{3}{2}$. Comme nous l'avons annoncé plus haut, c'est la valeur prise par le seul ratio $\dfrac{p_1}{p_2}$ qui conditionne le sens de la demande excédentaire. Ce ratio est précisément ce que l'on désigne sous l'appellation « termes de l'échange ». Comment le lire ? Un rapport d'échange $\dfrac{p_1}{p_2}$ égal à $\dfrac{3}{2}$ signifie que les protagonistes échangent 3 unités de bien 2 (soldats) contre 2 unités de biens 1 (autos) (ce qui est équivalent à 1 soldat contre 0,66 auto). Si ce rapport d'échange devient plus faible (par exemple 0,66 contre 0,66, c'est-à-dire 1 contre 1), Anatole aura une demande excédentaire en bien 1 (autos) positive, c'est-à-dire le souhait d'obtenir plus d'autos en cédant des soldats. Par un calcul analogue, il est facile d'établir d'autre part que $z_2^{\,A} \geq 0 \Leftrightarrow \dfrac{p_1}{p_2} \geq \dfrac{3}{2}$.

Ce constat est cohérent avec le précédent : si le rapport d'échange envisagé devient supérieur à $\dfrac{3}{2}$ (par exemple 2 soldats contre 1 auto), Anatole aura une demande excédentaire en soldats positive, c'est à dire le souhait de céder des autos pour obtenir des soldats supplémentaires. Ce résultat n'est qu'une autre manière d'exprimer le constat graphique que nous avions fait : la valeur absolue du *TmS* d'Anatole au point de ses dotations initiales est précisément égale à $\dfrac{3}{2}$: c'est la valeur seuil en dessous de laquelle sa demande excédentaire en bien 1 est positive et au-dessus de laquelle c'est sa demande excédentaire en bien 2 qui l'est.

5

Si nous examinons maintenant les demandes excédentaires de Brice, nous obtenons que : $z_1^B \geq 0 \Leftrightarrow \frac{2p_2 - 6p_1}{2p_1} \geq 0 \Leftrightarrow 2p_2 - 6p_1 \geq 0 \Leftrightarrow 2p_2 \geq 6p_1$

$\Leftrightarrow \frac{p_1}{p_2} \leq \frac{1}{3}$ et réciproquement que sa demande excédentaire en bien 2 est positive lorsque $\frac{p_1}{p_2} \geq \frac{1}{3}$. En précisant qu'une **demande excédentaire négative** correspond à une **offre excédentaire** (positive) du bien, il est possible de résumer ces résultats dans le tableau 1.

Tableau 1 Offres et demandes excédentaires en autos selon les termes de l'échange

Termes de l'échange	$\frac{p_1}{p_2} \leq \frac{1}{3}$	$\frac{1}{3} \leq \frac{p_1}{p_2} \leq \frac{3}{2}$	$\frac{3}{2} \leq \frac{p_1}{p_2}$
Anatole	Demande excédentaire d'autos	Demande excédentaire d'autos	Offre excédentaire d'autos
Brice	Demande excédentaire d'autos	Offre excédentaire d'autos	Offre excédentaire d'autos

De manière très évidente, lorsque les termes de l'échange $\frac{p_1}{p_2}$ sont compris entre $\frac{1}{3}$ et $\frac{3}{2}$, Anatole et Brice vont trouver un intérêt à l'échange car Anatole exprime une demande excédentaire d'autos tandis que Brice exprime une offre excédentaire de ce même bien. L'échange est alors possible car mutuellement désiré et mutuellement bénéfique. Remarquons qu'il s'agit bien sûr de la même conclusion que celle mise en évidence dans la section 1 où nous avions établi que le rapport d'échange devait se situer entre 1 (soldat) contre 3 (autos) et 1 (soldat) contre 0,66 (auto) (puisque $\frac{3}{2} = \frac{1}{0{,}66}$).

Mais grâce à l'expression des demandes excédentaires, nous pouvons aller plus loin, et établir avec précision le rapport d'échange « d'équilibre », c'est-à-dire celui conduisant à une allocation vers laquelle les deux enfants seront satisfaits de cheminer et de laquelle, lorsqu'elle sera atteinte, aucun ne souhaitera s'écarter.

L'équilibre général d'une économie d'échange est une allocation telle que, si les agents de cette économie l'ont atteinte, ils ne s'en écarteront pas.

Le rapport d'échange d'équilibre se détermine en résolvant un système d'équations qui établit que la somme des demandes excédentaires relatives à chaque bien, pris séparément, doit être nulle. Dans le cas d'une économie constituée de seulement deux agents (comme c'est la cas ici), ceci revient à exiger que l'offre excédentaire de l'un soit égale à la demande excédentaire de l'autre. Nous voici face à une forme qui nous est plus familière et que nous étudierons de manière détaillée dans le chapitre 8 : l'équilibre d'un marché est atteint quand l'offre est égale à la demande sur ce marché.

Résolvons maintenant ce système d'équations :

$$\begin{cases} z_1^A + z_1^B = 0 \\ z_2^A + z_2^B = 0 \end{cases} \Leftrightarrow \begin{cases} \dfrac{6p_2 - 4p_1}{3p_1} + \dfrac{2p_2 - 6p_1}{2p_1} = 0 \\ \dfrac{4p_1 - 6p_2}{3p_2} + \dfrac{6p_1 - 2p_2}{2p_2} = 0 \end{cases} \Leftrightarrow$$

$$\begin{cases} 9p_2 - 13p_1 = 0 \\ 13p_1 - 9p_2 = 0 \end{cases} \Leftrightarrow \begin{cases} \dfrac{p_1}{p_2} = \dfrac{9}{13} \\ \dfrac{p_1}{p_2} = \dfrac{9}{13} \end{cases}$$

D'un point de vue mathématique, ce système est « sous-déterminé ». D'un point de vue économique, nous pouvons conclure que jamais nous ne pourrons établir distinctement un prix p_1 et un prix p_2 : seul un prix relatif $\frac{p_1}{p_2}$, les termes de l'échange, pourra être obtenu.

Nous constatons que ce rapport d'échange d'équilibre $\frac{9}{13} \approx 0{,}692$ est effectivement compris entre $\frac{1}{3} = 0{,}333$ et $\frac{3}{2} = 1{,}5$. Il nous reste donc simplement à déterminer la position de l'allocation d'équilibre *E* dans la boîte d'Edgeworth.

En toute logique, nous pouvons établir que cette allocation sera une allocation Pareto-optimale. Si tel n'était pas le cas, les agents ne souhaiteraient pas demeurer en cette allocation car un nouvel échange serait susceptible d'accroître simultanément leur satisfaction.

Nous pouvons donc affirmer qu'au point *E*, $TmS_{1/2}{}^{A} = TmS_{1/2}{}^{B}$.

En outre, ces deux *TmS* sont aussi égaux au rapport d'échange d'équilibre puisqu'en ce point Anatole comme Brice réalisent leur décision optimale de consommation (caractérisée par l'égalité entre le *TmS* et le rapport des prix).

Enfin, ce point *E* doit appartenir à la boîte d'Edgeworth, ce qui signifie, dans notre application, que nécessairement $x_1{}^{A} + x_1{}^{B} = 10$ et $x_2{}^{A} + x_2{}^{B} = 5$.

Techniquement, il s'agit donc de résoudre le système d'équations :

$$\begin{cases} TmS^{A}_{1/2} = -\dfrac{p_1}{p_2} \\ TmS^{B}_{1/2} = -\dfrac{p_1}{p_2} \\ x_1^A + x_1^B = 10 \\ x_2^A + x_2^B = 5 \end{cases} \Leftrightarrow \begin{cases} \dfrac{2x_2^A}{x_1^A} = \dfrac{9}{13} \\ \dfrac{x_2^B}{x_1^B} = \dfrac{9}{13} \\ x_1^A + x_1^B = 10 \\ x_2^A + x_2^B = 5 \end{cases}$$

La résolution de ce système donne :
$x_1{}^{A} = \dfrac{50}{9} \approx 5{,}56$ $x_2{}^{A} = \dfrac{25}{13} \approx 1{,}92$ $x_1{}^{B} = \dfrac{40}{9} \approx 4{,}44$ et $x_2{}^{A} = \dfrac{40}{13} \approx 3{,}08$. Ce sont les coordonnées de l'allocation d'équilibre *E*.

En d'autres termes, Anatole a cédé 3 – 1,92 = 1,08 unités de bien 2 pour obtenir 5,56 – 4 = 1,56 unités supplémentaire de bien 1. Symétriquement, Brice a obtenu 1,08 unités de bien 2 en cédant 1,56 unités de bien 1. Bien entendu $\dfrac{1{,}08}{1{,}56} = \dfrac{9}{13} \approx 0{,}692$. Ce cheminement vers l'équilibre général est représenté figure 8.

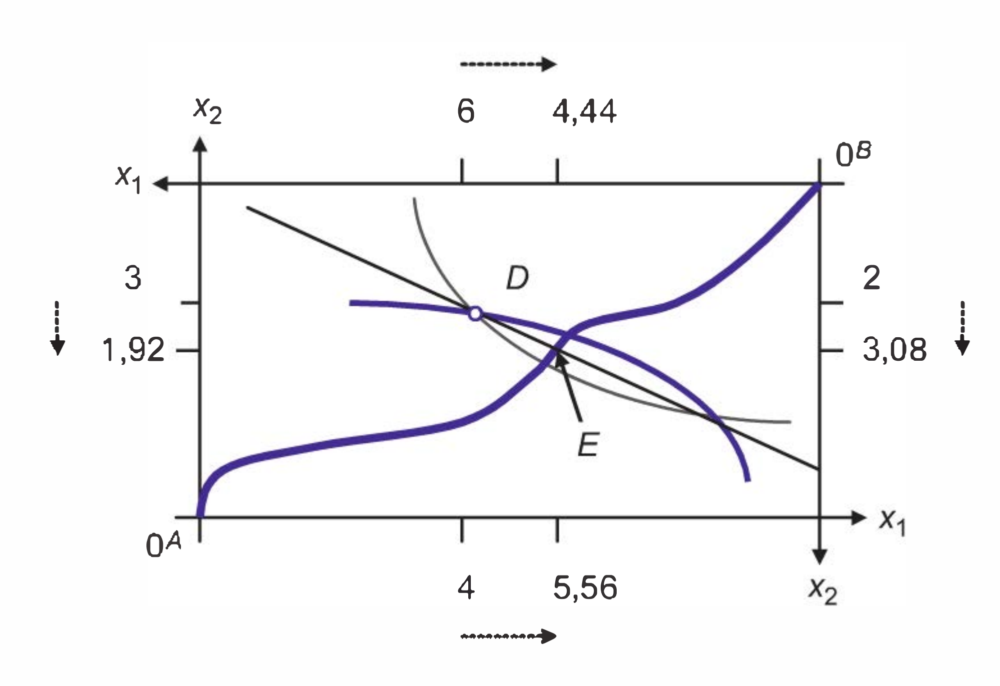


Figure 8 Équilibre général de l'économie d'échange

Le point E est l'équilibre général de cette économie d'échanges. L'équilibre est atteint suite à un processus de troc spontané entre les deux consommateurs selon un rapport d'échanges qui se détermine par la confrontation de leurs désirs d'échanges. Si l'on est peu convaincu par l'idée que le processus de troc conduise « en une seule fois » à l'équilibre, on peut envisager un processus avec tâtonnement. C'est l'idée qu'a originellement suggérée le père du concept d'équilibre général, Léon Walras. Il a formalisé cette idée en évoquant la présence d'un commissaire-priseur (fictif) proposant des prix (des rapports d'échange) aux différents agents, et les révisant, en fonction des intentions d'offre et de demande, jusqu'à égalisation de toutes les offres et toutes les demandes. Ce processus de tâtonnement dit walrassien conduit à un système de prix d'équilibre, sous-entendu équilibre de tous les marchés simultanément ou équilibre général.

L'une des caractéristiques de l'équilibre walrassien est qu'il suffit que $(n-1)$ marchés soient en équilibre pour que le $n^{\text{ième}}$ le soit aussi. En conséquence, on ne peut déterminer que $(n-1)$ prix relatifs et non n prix absolus. L'une des lectures que l'on peut adopter est de considérer que la monnaie est le $n^{\text{ième}}$ bien, le bien numéraire, celui dont le prix est égal à 1. Ainsi les $(n-1)$ autres prix deviennent des prix absolus au sens usuel du terme.

La construction théorique que nous avons ébauchée dans ce chapitre débouche sur l'un des résultats centraux de la théorie économique : le **premier théorème de l'économie du bien-être**.

Théorème

Dans une économie de concurrence pure et parfaite, sans externalités, si les préférences des consommateurs sont strictement convexes, l'équilibre général est un optimum de Pareto.

Plusieurs des concepts évoqués dans cette proposition (concurrence pure et parfaite, externalités) n'ont pas encore été étudiés et le seront dans les chapitres suivants. Nous pouvons néanmoins retenir le message fort de ce théorème. Si, lorsque certaines conditions sont réunies, on laisse les agents produire, échanger et consommer comme bon leur semble, on aboutit à une allocation des ressources optimales au sens de Pareto (sans intervention extérieure, en particulier celle de l'État). L'enseignement de ce théorème peut être qualifié de « libéral ». Il s'agit de l'une des pierres angulaires du courant néo-classique, le courant dominant de la théorie économique. Il convient néanmoins de nuancer cette lecture « libérale ». En effet :

- Dès lors que les préférences des consommateurs ne sont pas toutes strictement convexes, ce théorème n'est pas vérifié. Rappelons qu'un consommateur ayant des préférences strictement concaves a un comportement addictif ou de collectionneur ; s'il dispose de grandes quantités du bien qu'il apprécie le plus (et de faibles quantités de l'autre) il sera très réticent à l'échange. En l'absence d'échange, l'économie demeure au point des dotations initiales. Sauf cas fortuit, cette allocation n'est pas optimale.
- Dès lors qu'il existe des biens publics (justice, ordre public, sécurité, éclairage public, parcs, etc.) et/ou des effets externes – encore appelés externalités – (pollution), l'allocation des ressources issue du jeu spontané des rapports économiques, n'est plus optimale au sens de Pareto. Ceci justifie l'intervention de la puissance publique.
- Dans une économie élargie à la production, il se peut que certains marchés aient un fonctionnement qui ne soit pas de type concurrentiel

(au sens de la concurrence pure et parfaite) : monopole, oligopole, etc. Dès lors, le théorème ci-dessus n'est plus vérifié.

- Enfin, comme nous l'avons déjà dit, l'efficacité économique au sens de Pareto ne signifie pas nécessairement que l'allocation des ressources sera juste. Là encore, une intervention de la puissance publique visant à atteindre une allocation des ressources plus « équitable » peut être justifiée. C'est le principe même de la redistribution, c'est-à-dire de la mise en œuvre de dispositifs permettant de verser des prestations sociales. Notons qu'au titre d'une « bonne » politique sociale, on s'efforcera néanmoins d'atteindre un autre optimum de Pareto sur la courbe des contrats.

6 La combinaison optimale des facteurs de production

Mots-clés

Technologie de production, input, output, productivité marginale, rendements d'échelle, isoquant, taux marginal de substitution technique, coût total de production, coût moyen, coût marginal

Quels sont les déterminants de l'offre des biens et services consommés par les individus dans une économie ? Nous étudions, dans ce chapitre, la manière dont un entrepreneur, désireux de produire un bien ou service, optimise son processus de production.

Par hypothèse, les caractéristiques du bien (ou de la prestation de service) dont il envisage d'organiser la production sont déjà parfaitement définies. Cette hypothèse n'est pas triviale car elle suppose que l'on puisse connaître parfaitement les goûts, en perpétuelle évolution, des consommateurs et que l'on sache répondre, de façon appropriée, à leurs désirs.

1 Formalisation de la technologie de production

Notre première tâche consiste à modéliser la manière dont il est possible de produire un certain nombre d'unités d'un bien ou service (dont les caractéristiques sont parfaitement définies) appelé **output**, à partir de la transformation d'un ou plusieurs facteurs de production appelés **inputs**.

Une **technologie de production** est un processus de transformation basé sur la combinaison de multiples facteurs de productions (nommés inputs) et conduisant à l'élaboration, la confection, la réalisation d'un bien ou service (nommé output).

1.1 Les facteurs de production ou inputs

Les « inputs », composants utilisés dans les processus de transformation, sont des éléments de natures très différentes. Il peut s'agir d'hectares de terre, de tonnes de matières premières, de kilowatts d'énergie, d'heures de travail, de temps d'utilisation de machines, de prestations de nettoyage, etc. Parmi tous ces inputs, il en est un que les économistes ont pris l'habitude de désigner par le terme « capital ».

Le capital

En microéconomie, le **capital** désigne des inputs qui sont eux-mêmes des biens produits. En d'autres termes, ces inputs sont des biens déjà issus d'un processus de transformation.

Les inputs regroupés sous la dénomination « capital » interviennent de deux manières dans la production de l'output :

- soit ils sont des « machines » utiles au processus de production (machines à tisser, robots à emboutir ou peindre les tôles, chariots élévateurs, véhicules, etc.) ;
- soit ils sont des ingrédients entrant dans la fabrication de l'output, pour certains qualifiés de « biens intermédiaires » (vis et boulons, plaques, blocs, pièces, en métal, en matière plastique, en bois, en résine, composants électroniques, etc.).

Les inputs « ingrédients » – entrant dans la fabrication de l'output – ont en commun d'être utilisés en quantités variables, en fonction de la quantité d'output que l'entrepreneur souhaite produire : on parle d'éléments de **capital variable**. À l'inverse, les machines utiles au processus de production sont acquises (ou louées) indépendamment des quantités d'output produites :

elles sont les éléments du **capital fixe**. La distinction entre capital fixe et capital variable est parfois ténue : si le crayon utilisé par le dessinateur pour produire des plans et croquis est incontestablement un élément de capital variable, il est plus difficile de situer la pelle utilisée par le terrassier pour produire du terrassement ou l'aspirateur utilisé par l'homme ou la femme de ménage pour produire du nettoyage. Si l'avion de ligne utilisé par une compagnie aérienne pour produire des trajets aériens est un élément de capital fixe, certaines de ses « pièces d'usure » (tels que les pneumatiques) sont des éléments de capital variable. Bien entendu, les immeubles destinés à la production (bureaux, ateliers, boutiques, etc.) sont eux aussi à inclure dans le capital fixe. Le capital est donc un input extrêmement hétérogène ce qui ne manque pas de susciter certaines critiques méthodologiques. Ces critiques sont amplifiées par la confusion fréquente accompagnant la distinction entre capital physique et capital financier.

Dans le langage courant, le terme capital désigne principalement, en effet, les sommes d'argent utilisées pour faire démarrer ou fonctionner une entreprise. Or, ce capital financier est utilisé à l'acquisition de tous types d'input (dont les matières premières et les heures de travail) et pas seulement à l'achat de capital au sens microéconomique du terme. Convenons donc que, dans toute la suite, le terme capital n'évoquera que le capital physique, c'est-à-dire l'ensemble des composants entrant dans le processus réel (par opposition à monétaire) de transformation.

■ Le travail

Le travail est, lui aussi, un input regroupant des éléments de natures et de qualités très variées. La variété des compétences et des aptitudes génère d'ailleurs une disparité importante des rémunérations. Dans les modèles d'économie du travail, on rencontre souvent une distinction entre les travailleurs qualifiés et les travailleurs non qualifiés, ou, mieux encore, une distribution continue des travailleurs selon leur niveau de qualification.

En dépit de la sophistication d'une telle modélisation, la réalité demeure plus complexe encore. En effet, certaines prestations sont uniquement intellectuelles, d'autres allient les talents du corps et de l'esprit, d'autres encore sont la simple et fastidieuse répétition de tâches identiques. De plus,

le travail en équipe est supposé engendrer des phénomènes vertueux que la seule mesure individuelle des talents ne permet pas de détecter.

Enfin, des contraintes institutionnelles fortes et nombreuses font du facteur travail un input très complexe, multidimensionnel, d'une disponibilité et d'une flexibilité imparfaites. On suppose néanmoins, dans nos modélisations, qu'à l'instar du facteur capital, le facteur travail est une variable unidimensionnelle convenablement définie, qui intervient comme ingrédient dans le processus de production de l'output analysé.

1.2 Des inputs à l'output

Modélisons maintenant la « technologie de production » ou processus par lequel, en combinant un ou plusieurs inputs, on aboutit à la production d'un certain nombre d'unités d'output.

Cas d'un seul facteur de production

Envisageons d'abord le cas où un seul input est nécessaire pour produire l'output. Le cas le plus courant est celui où seul l'input travail est utilisé. Par exemple, un chanteur des rues produit des chansons à l'aide de sa seule voix ou un cueilleur de mûres sauvages produit une récolte à l'aide de ses seules mains. Dans les deux cas, l'input unique est bien le travail, mesuré en heures.

Représentons sur la figure 1, le nombre de kilogrammes de mûres cueillies en fonction du nombre d'heures consacrées à la récolte par un cueilleur. Nous choisissons de désigner par Q le nombre de kilogrammes de mûres sauvages cueillies et par x_1 le nombre d'heures de travail consacrées à la cueillette.

Sur cette courbe, nous pouvons « lire » que si le cueilleur travaille 1 heure, il sera en mesure de cueillir environ 4,5 kilogrammes de mûres et que s'il travaille une heure de plus, la récolte totale atteindra environ 6,5 kilogrammes ; en d'autres termes, il ne sera parvenu à cueillir que 2 kgs supplémentaires lors de la seconde heure de travail. On peut expliquer un tel phénomène par la raréfaction des mûres dans les haies ou en prétextant la fatigue du cueilleur qui, après avoir été très performant pendant la première heure de cueillette, est victime d'une « baisse de régime » pendant la seconde heure.

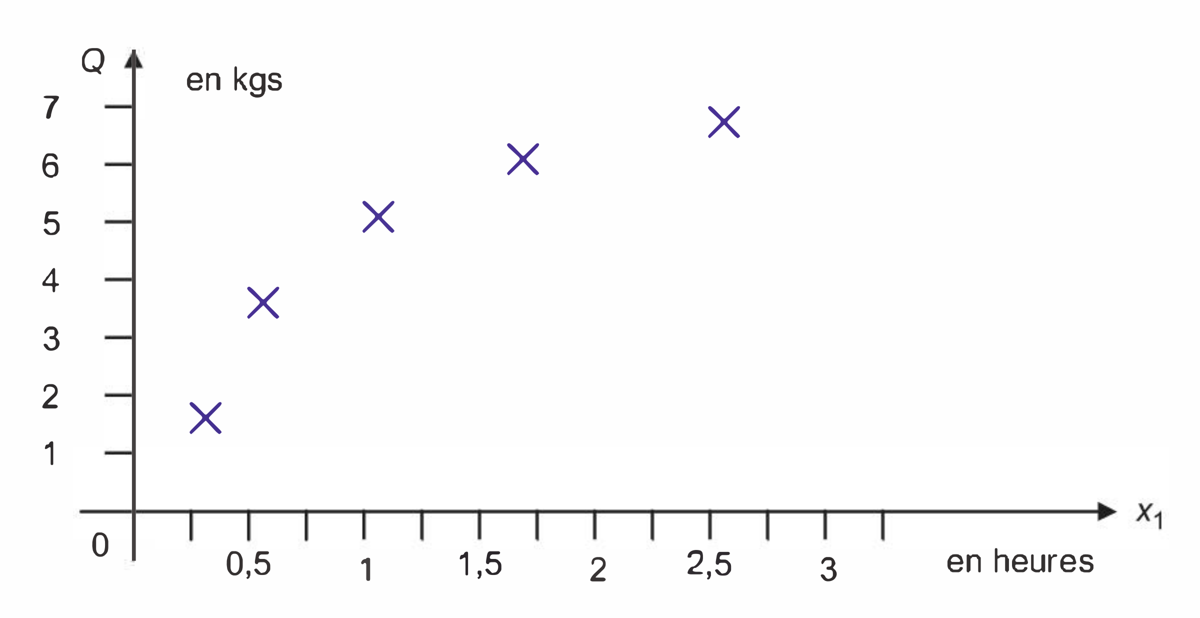


Figure 1 Nombre d'unités d'output produites en fonction du nombre d'heures travaillées

Ce que nous voyons s'ébaucher est bien une courbe qui représente la relation fonctionnelle entre nombre d'heures de travail (input) utilisées et nombre de kilogrammes de mûres (output) cueillies. Plus précisément, la courbe apparaissant sur la figure 2 indique le nombre **maximal** d'unités d'output qu'il est possible d'obtenir pour toute quantité utilisée d'input.

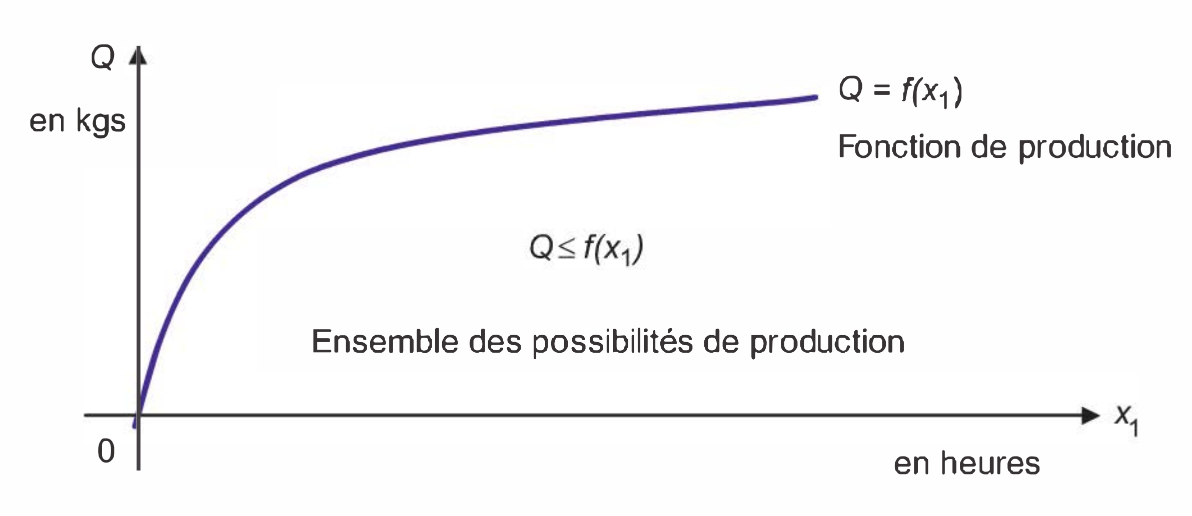


Figure 2 Fonction de production de l'output (cas d'un seul input)

La courbe est la représentation graphique de la relation fonctionnelle (la fonction) $Q = f(x_1)$, que l'on nomme « fonction de production ». Si un

facteur est gaspillé (le cueilleur interrompt sa cueillette et utilise une heure pour s'adonner à la sieste) ou si l'on prélève une partie de la production (dans notre exemple, par gourmandise) la quantité produite est inférieure à la quantité maximale figurant sur la fonction de production. Pour cette raison, il est opportun de spécifier la technologie de production comme : $Q \leq f(x_1)$. Géométriquement, cette inégalité délimite toute la surface située sous la courbe, que l'on nomme « ensemble des possibilités de production ». De manière générale, la présence d'une inégalité permet de délimiter, dans la formalisation, ce qui **accessible** en matière de quantités produites.

La fonction de production présentée sur la figure 2 est croissante à taux décroissant. Le fait qu'elle soit croissante est intuitif : plus on consacre de temps à la cueillette, plus on obtient de mûres. Le fait que cette croissance se fasse à taux décroissant est la traduction de ce que le cueilleur va, pour chaque heure supplémentaire travaillée, cueillir une quantité supplémentaire de mûres de plus en plus faible (fatigue et/ou raréfaction de la ressource). Dans d'autres circonstances, une fonction de production pourrait être croissante à taux constant. En particulier, si l'input en jeu est du capital (variable), il se peut que chaque unité supplémentaire d'input utilisée permette de produire toujours la même quantité supplémentaire d'output : par exemple, s'il faut 400 grammes de granulés de plastique pour produire une corbeille à papier, 800 grammes permettront d'en produire deux, 1 200 grammes permettront d'en produire trois, et ainsi de suite… Cet exemple demeure un peu artificiel car il est difficile de produire un tel output à partir d'un input unique. C'est pourquoi il faut élargir notre perspective et envisager une technologie de production où les unités d'output sont obtenues à partir de la combinaison de deux ou plusieurs inputs.

Cas de deux facteurs de production

Supposons maintenant que l'output est produit à partir de deux facteurs. Nous notons Q la quantité de bien produit (output), x_1 la quantité utilisée du premier facteur (par exemple le travail), x_2 la quantité utilisée du second facteur (par exemple le capital). Représentons, sur la figure 3, la technologie de production.

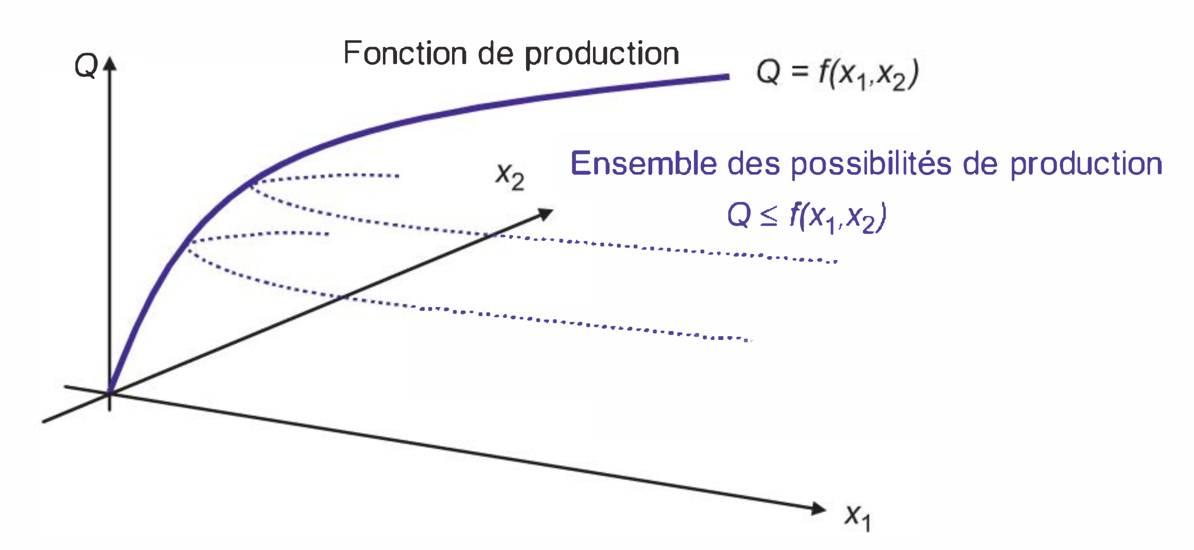


Figure 3 Fonction de production de l'output (cas de deux inputs)

La relation fonctionnelle $Q = f(x_1 ; x_2)$ indique la quantité maximale d'output que l'on peut produire avec une quantité donnée d'input. Comme précédemment, la technologie de production est exprimée par l'inégalité $Q \leq f(x_1 ; x_2)$: cela signifie qu'il est possible de produire moins que la quantité maximale d'output si on gaspille les inputs, si on les combine mal ou si on perd ou détériore les outputs pendant leur production (par exemple si on laisse des produits frais, juste transformés, dans un lieu de stockage inapproprié…). Bien sûr, on s'efforce d'éviter tout gaspillage et on tente de se positionner sur la fonction de production elle-même, parfois encore appelée « enveloppe supérieure de l'ensemble des possibilités de production ».

On peut généraliser ce raisonnement à une technologie de production faisant intervenir n inputs, dont les quantités utilisées sont désignées par x_1, x_2, …, x_n. L'ensemble des possibilités de production est alors le domaine délimité par l'inégalité $Q \leq f(x_1 ; x_2 ; \ldots ; x_n)$ et la fonction de production est la fonction de n variables $Q = f(x_1 ; x_2 ; \ldots ; x_n)$.

On appelle **fonction de production** d'un output produit en quantité Q à partir de n inputs utilisés en quantités x_1, x_2, …, x_n la relation fonctionnelle qui décrit la quantité maximale d'output qu'il est possible de produire à partir de la combinaison de ces inputs. On note cette fonction de production : $Q = f(x_1 ; x_2 ; \ldots ; x_n)$.

6

L'inégalité $Q \leq f(x_1 ; x_2 ; \ldots ; x_n)$ délimite l'ensemble des possibilités de production, c'est-à-dire l'ensemble des combinaisons d'inputs permettant d'atteindre une certaine cible de production Q sans que la combinaison utilisée ne soit nécessairement la plus adaptée ou la plus efficace. Seule une combinaison réalisant l'égalité $Q = f(x_1 ; x_2 ; \ldots ; x_n)$ sera, du point de vue de la firme, pleinement satisfaisante.

2 Productivité marginale et rendements d'échelle

Dès lors que nous abordons la question de la production d'output à partir de la combinaison de deux ou plusieurs inputs, il est utile de construire des outils qui nous permettront de caractériser les propriétés de la technologie de production :

- Le surcroît de production induit par l'utilisation d'une quantité supplémentaire d'un input particulier (isolément) est-il moins que proportionnel, exactement proportionnel, ou plus que proportionnel ?
- Et si nous augmentons la quantité utilisée de tous les inputs dans les mêmes proportions, quel en est l'impact sur la quantité totale produite ?

Deux outils permettent de traiter ces questions. Ces outils sont les **productivités marginales des facteurs** et les **rendements d'échelle**. En étudiant si la productivité marginale d'un facteur est décroissante, constante (ou même croissante), on établit comment se comporte la technologie de production suite à l'accroissement de la quantité utilisée d'un facteur pris isolément ; en étudiant si les rendements d'échelle sont décroissants, constants ou croissants, on établit comment se comporte la technologie suite à l'accroissement de la quantité utilisée de tous les facteurs simultanément et dans les mêmes proportions.

2.1 Productivité marginale

Soit $f(x_1 ; x_2 ; \ldots ; x_n)$ une fonction de production. On appelle **productivité marginale du facteur** i, notée $Pm_i(x_1 ; x_2 ; \ldots ; x_n)$ le surcroît de production induit par un accroissement infinitésimal de la quantité utilisée du facteur *i*. Formellement, la productivité marginale du facteur *i* est la dérivée partielle de la fonction de production relativement à la variable x_i :

$$Pm_i(x_1 ; x_2 ; \ldots ; x_n) = \frac{\partial f(x_1 ; x_2 ; \ldots ; x_n)}{\partial x_i}$$

Souvent, pour simplifier, on dit que la productivité marginale d'un facteur est le surcroît de production induit par un accroissement de 1 % de la quantité utilisée de l'input. Ainsi, par exemple :

- si la production s'accroît de 0,8 %, suite à un accroissement de 1 % de la quantité utilisée du facteur travail, on dit que la productivité marginale du travail est décroissante ;
- si la production s'accroît de 1 %, suite à un accroissement de 1 % de la quantité utilisée du facteur travail, on dit que la productivité marginale du travail est constante.

6

2.2 Rendements d'échelle

Soit $f(x_1 ; x_2 ; \ldots ; x_n)$ une fonction de production. On dira que les **rendements d'échelle** associés à cette fonction de production sont croissants (respectivement décroissants, constants) si le surcroît de production induit par un accroissement infinitésimal, de même ampleur, de tous les facteurs simultanément, est plus que proportionnel (respectivement proportionnel, moins que proportionnel).

Pour juger du caractère croissant, constant ou décroissant des rendements d'échelle, on calcule l'impact, sur la quantité d'output produit, de la multiplication de tous les x_i par un même nombre réel $\lambda > 1$:

- Si $f(\lambda x_1 ; \lambda x_2 ; \ldots ; \lambda x_n) > \lambda \ f(x_1 ; x_2 ; \ldots ; x_n)$, les rendements d'échelle sont croissants.

- Si $f(\lambda x_1\ ;\ \lambda x_2\ ;\ \ldots\ ;\ \lambda x_n) = \lambda\ f(x_1\ ;\ x_2\ ;\ \ldots\ ;\ x_n)$, les rendements d'échelle sont constants.
- Si $f(\lambda x_1\ ;\ \lambda x_2\ ;\ \ldots\ ;\ \lambda x_n) < \lambda\ f(x_1\ ;\ x_2\ ;\ \ldots\ ;\ x_n)$, les rendements d'échelle sont décroissants.

Concrètement, pour étudier l'impact d'un accroissement infinitésimal de la quantité utilisée de tous les facteurs simultanément (et dans les mêmes proportions), on peut s'intéresser à un accroissement de 1 % de la quantité utilisée des inputs en appliquant la valeur $\lambda = 1{,}01$.

Une technologie de production peut-elle être caractérisée à la fois par des productivités marginales décroissantes et des rendements d'échelle croissants ? En d'autres termes, même s'il est raisonnable de supposer que la productivité marginale de tous les inputs (pris isolément) est, au mieux, constante (les phénomènes physiques sont tous sujets à des frottements et autres sources de déperdition), est-il déraisonnable de concevoir que l'alchimie de la combinaison des inputs puisse conduire à une croissance de la quantité produite plus que proportionnelle (aux augmentations des quantités utilisées des différents facteurs) ?

Posons le problème dans un cas imaginaire : supposons qu'une ligne de production fasse intervenir 25 opérateurs sur 5 machines. Il est concevable que l'accroissement de production induit par la présence de 5 opérateurs supplémentaires (+ 20 %) (en conservant une ligne de 5 machines) soit inférieur à 20 %. Il est également concevable que l'accroissement de production induit par la présence d'1 machine supplémentaire (+ 20 %) (en n'accroissant pas le nombre d'opérateurs) soit aussi inférieur à 20 %. Il est cependant possible que l'accroissement de production induit par la présence simultanée de 5 opérateurs supplémentaires et d'1 machine supplémentaire soit supérieur à 20 %, en raison, par exemple, d'une meilleure division des tâches rendue possible par la multiplication des combinaisons possibles de séquences de transformation et de montage. En résumé, des rendements d'échelle croissants sont possibles. Notons que ce qui est décrit ici fait référence à des facteurs utilisés en quantité variable : même si, dans l'exemple ci-dessus, on évoque la présence de « machines » de production, leur nombre fait ici implicitement l'objet d'un ajustement sans délai, si le besoin s'en fait sentir.

Si, à l'inverse, il s'agissait de facteurs disponibles en quantité fixe (terrain, bâtiments, gros équipements, etc.) et dont la capacité ou le nombre ne peut s'ajuster qu'à l'issue d'un long délai (plusieurs années), il serait plus difficile de se prononcer sur la question des rendements d'échelle tels que nous les avons définis ici.

Explorons maintenant la question de la substituabilité entre les facteurs : peut-on remplacer facilement du travail par des machines (ou vice-versa) dans le cadre de la production d'un output, et si oui, dans quelles proportions ? L'étude graphique des courbes d'iso-quantités d'output (ou isoquants) va nous aider à répondre à ces questions.

3 Représentation graphique sous forme d'isoquants

Dans le cas où la technologie de production ne nécessite l'utilisation que de deux inputs, il est possible de représenter graphiquement, de manière simple, les ensembles d'iso-production d'output, ou isoquants.

3.1 Isoquants

Soit $Q = f(x_1 ; x_2)$ une fonction de production. Un **isoquant** est un ensemble de combinaisons des inputs 1 et 2, c'est-à-dire de couples $(x_1 ; x_2)$, pour lesquels il est possible d'atteindre, au mieux, une production Q_0. Pour une telle cible de production Q_0, l'équation de l'isoquant sera : $Q_0 = f(x_1 ; x_2)$.

Dans le repère $(0 ; x_1 ; x_2)$, l'isoquant d'équation $Q_0 = f(x_1 ; x_2)$ prend la forme d'une courbe de niveau, à l'image de ce que nous avions pour les courbes d'indifférence dans les chapitres précédents : chaque courbe représente désormais un niveau donné de la quantité produite. Représentons sur la figure 4, l'isoquant relatif à un niveau de production particulier Q_0.

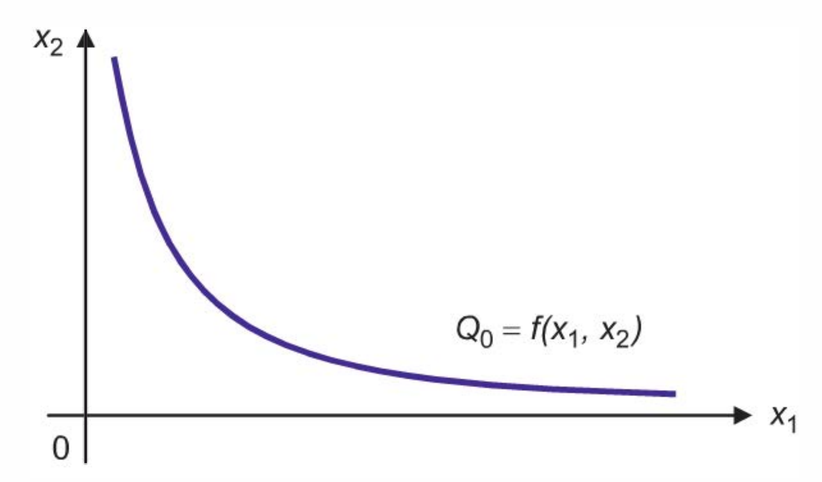


Figure 4 Isoquant relatif à un niveau de production Q_0

L'isoquant apparaissant sur la figure 4 est continu, dérivable et convexe (car toute combinaison linéaire convexe de points de l'isoquant appartient à un ensemble qui est lui-même convexe, cf. chapitre 1). La dérivabilité (propriété se concrétisant par l'absence de point anguleux) et plus encore la continuité des isoquants sont des hypothèses très peu réalistes. Dans le meilleur des cas, dans la réalité, les isoquants seraient une collection finie de points. Ceci est dû, en particulier, à l'indivisibilité de certains inputs (le travail est relativement divisible, les machines nettement moins). Pire encore, suggérer, comme nous le faisons, que les différents points de l'isoquant sont différentes manières de produire une certaine quantité d'output à partir d'une même technologie, technologie pour laquelle il serait possible de substituer continûment un facteur par l'autre, est très éloigné de la réalité. La production des biens est souvent, telle une recette de cuisine, la combinaison d'ingrédients dans des proportions très précises, sans que l'on puisse réellement s'écarter de ces proportions. Si, dans la confection d'un gâteau, on décide de mettre plus de farine et moins de beurre, certes une modification infinitésimale des quantités n'aura pas grande incidence, mais, à partir d'un certain degré de substitution d'un ingrédient par l'autre, le résultat sera calamiteux et la production (le gâteau) pourra être jetée à la poubelle. Dans le cadre d'une production industrielle, on pourra, dans certains cas, envisager un petit nombre de techniques de production alternatives, mais alors il ne s'agit plus, formellement, d'une unique fonction de production à partir de laquelle il serait possible de calculer des taux de substitution entre les inputs comme nous allons nous y adonner ci-dessous. En résumé, l'ensemble des développements que nous

allons maintenant présenter doivent être considérés pour ce qu'ils sont : une représentation (très) simplifiée de la manière dont un entrepreneur tente de déterminer la combinaison optimale des facteurs de production.

Avant d'écrire le programme qui permettra de modéliser cette prise de décision, précisons une propriété centrale des isoquants.

Le niveau de production atteint est d'autant plus élevé que les isoquants sont situés vers le haut, vers la droite

Sur la figure 5, sont représentés deux isoquants Q_0 et Q_0'. Les quantités produites Q_0 et Q_0' sont telles que $Q_0' > Q_0$.

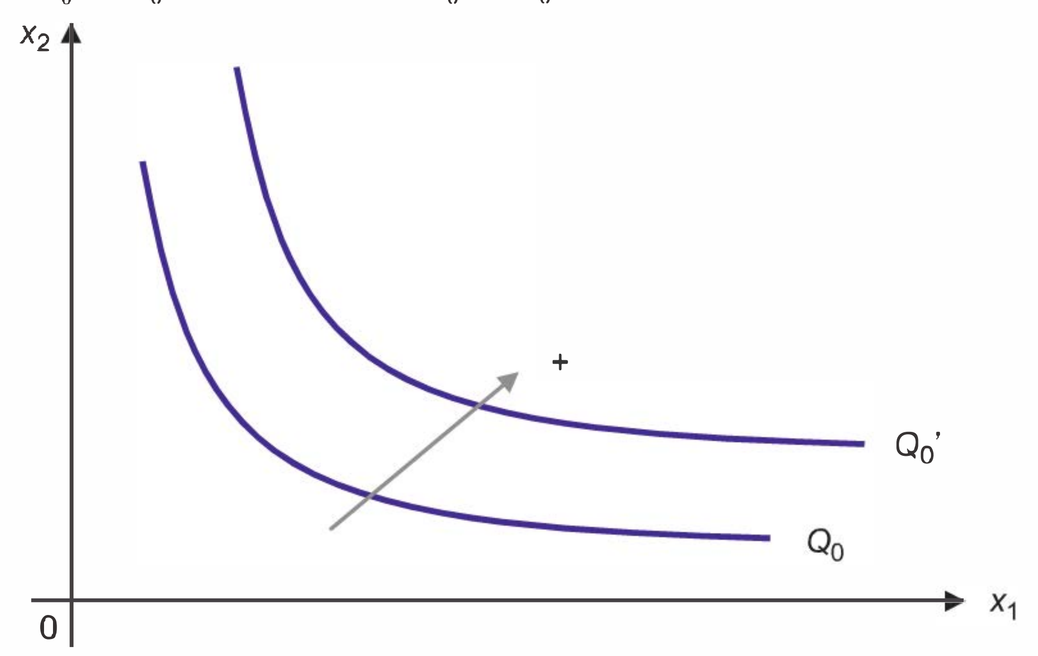


Figure 5 La quantité produite est d'autant plus grande que l'isoquant est situé vers le haut et la droite

Ceci est directement lié à l'hypothèse selon laquelle la production croît dès lors que l'on accroît la quantité utilisée de l'un quelconque des inputs. Mais cette hypothèse n'est pas toujours vérifiée. Supposons que le facteur 1 soit du travail (d'un agriculteur) et que le facteur 2 soit du terrain agricole. L'output est une céréale quelconque (du blé, par exemple). Jusqu'à une certaine limite, pour une surface donnée de terrain agricole (en ordonnée), plus l'agriculteur travaille longtemps (en abscisse), plus il exploite une fraction importante du terrain et plus il récolte une grande quantité de blé. Ceci n'est valable que tant qu'il n'exploite pas la surface totale du terrain.

Dès que la surface est complètement occupée, l'accroissement du nombre d'heures travaillées ne permet plus d'accroître la production. Un tel phénomène se rencontre dans la production de nombreux outputs. Ce type de circonstances s'apparente à (ou fait tendre vers) une technologie de production *à facteurs combinés en proportions fixes.*

3.2 Cas de facteurs combinés en proportions fixes (substituabilité nulle)

Supposons que pour produire 1 unité d'output, il faille des proportions fixes de deux facteurs. Par exemple, pour abattre des arbres, il faut équiper chaque bûcheron d'une et unique tronçonneuse (le fait de disposer de deux tronçonneuses par bûcheron ne permet pas d'abattre plus d'arbres par unité de temps ; de même si deux bûcherons se disputent une unique tronçonneuse, la quantité d'arbres abattus ne sera pas supérieure à la quantité atteinte par un travailleur seul). Représentons de tels isoquants sur la figure 6 (où $Q_0' > Q_0$).

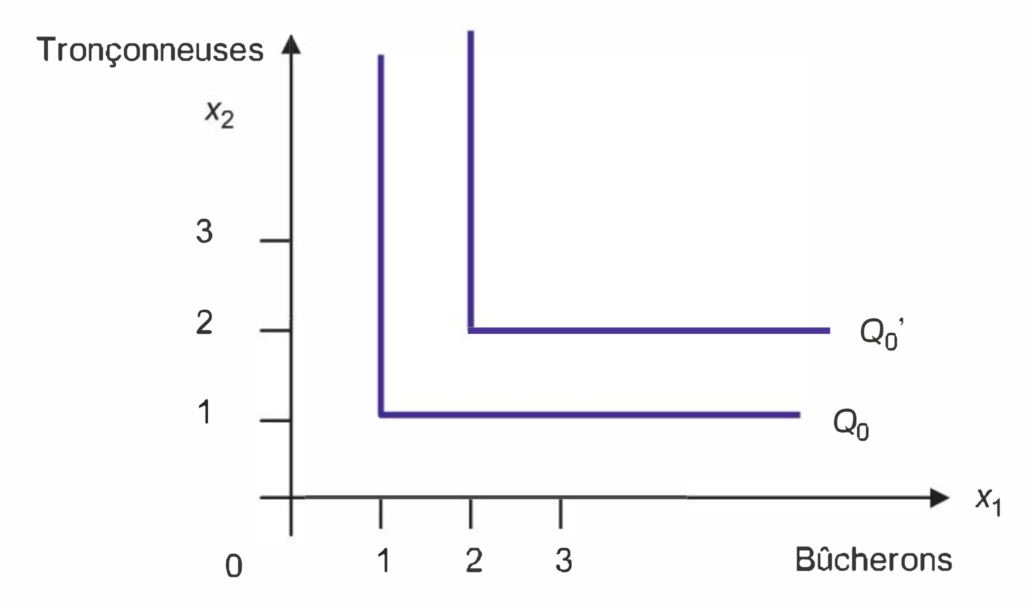


Figure 6 Isoquants dans le cas de facteurs combinés en proportions fixes

Comme dans le cas de courbes d'indifférences relatives à des biens complémentaires, les isoquants relatifs à des technologies où les facteurs sont combinés en proportions fixes prennent la forme de coudes. Toute comme

le niveau de satisfaction éprouvé lors de la consommation de biens complémentaires était « tiré vers le bas » par le nombre d'unités disponibles du bien présent en quantité la plus faible, la quantité d'output produite sera conditionnée par le nombre d'unités disponibles de l'input présent en quantité la plus faible. De tels isoquants sont donc obtenus à partir d'une fonction de production dont l'équation est de type : $Q = Min\{x_1 ; x_2\}$. Plus généralement, l'équation de fonctions de production mettant en œuvre des inputs combinés en proportions fixes est : $Q = Min\{ax_1 + b ; cx_2 + d\}$ où a, b, c et d désignent des nombres réels ($a, b > 0$).

3.3 Cas de facteurs parfaitement substituables

Prolongeons l'analogie entre les courbes d'indifférence et les isoquants en évoquant maintenant des technologies qui utiliseraient des facteurs parfaitement substituables. L'existence d'inputs parfaitement substituables est un peu hypothétique. Les entreprises, dans leur activité réelle, sont parfois dans l'obligation de trouver un input qui puisse remplacer un facteur précédemment utilisé (par exemple un ingrédient qui n'est plus acceptable au regard de nouvelles normes toxicologiques). Dans ce cas, l'entreprise s'efforce de trouver un input substitut en espérant que son utilisation ne modifiera pas (trop) la texture, le goût ou les propriétés de l'output qu'elle produit. Dans l'exemple que nous allons considérer, les deux facteurs substituts pourront être indifféremment et simultanément utilisés. Pour qu'un tel exemple illustratif puisse être construit, il faut nécessairement que les deux inputs considérés soient très faiblement différenciés (en réalité, même, la différence qui leur est prêtée est largement artificielle). Supposons que l'output considéré soit la livraison de pizzas.

Pour livrer des pizzas, on peut soit utiliser des scooters à moteurs électriques (facteur 1) soit des scooters à moteurs thermiques (facteur 2). Sur la figure 7, on représente l'isoquant correspondant à la livraison de $Q_0 = 30$ pizzas : cette prestation peut être opérée en livrant 20 pizzas par des scooters électriques et 10 par des scooters thermiques, ou 30 par des scooters électriques et 0 par des scooters thermiques, ou 10 par des scooters électriques et 20 par des scooters thermiques…

6

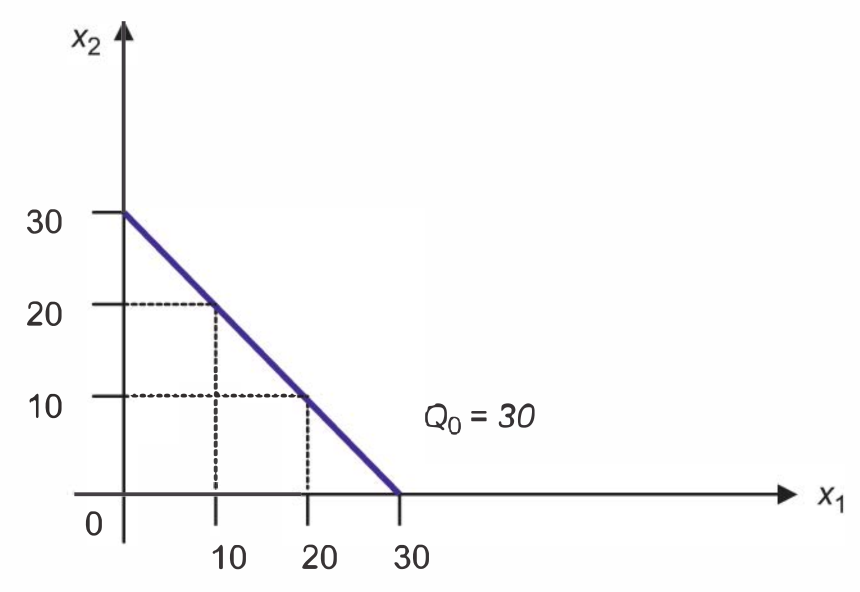


Figure 7 Isoquants dans le cas de facteurs parfaitement substituables

Le nombre d'unités d'output produites (le nombre de pizzas livrées) sera concrètement déterminé par le nombre total de trajets réalisés, et ceci indépendamment de la répartition entre ceux effectués à partir d'un mode de propulsion électrique ou à partir d'un mode de propulsion thermique.

Entre les deux cas extrêmes (facteurs combinés en proportions fixes et facteurs parfaitement substituables), il existe une variété considérable de situations. Nous n'allons pas construire une typologie des fonctions de production sur la base du degré de substituabilité des inputs car cette modélisation des technologies demeure assez lointaine de la réalité concrète au sein des entreprises. Néanmoins, l'intuition selon laquelle les facteurs sont, dans une certaine mesure, substituables est pertinente et nous allons maintenant détailler la manière de quantifier les proportions dans lesquels il est possible de substituer un input par un autre dans le cadre de la modélisation retenue. Ceci correspond à la notion de taux (marginal) de substitution technique ($TmST$).

Le taux marginal de substitution technique entre deux facteurs de production est une mesure des proportions dans lesquelles il est possible de substituer un input par un autre sans que la quantité d'output produite ne soit modifiée. On notera $TmST_{i/j}$ le taux marginal de substitution technique entre les inputs i et j.

Pour mieux cerner cette notion, aidons-nous d'arguments graphiques. Raisonnons sur un output produit à partir de deux facteurs selon la technologie de production $Q = f(x_1, x_2)$. Considérons une combinaison particulière des inputs, désignée par *A*, permettant d'atteindre une quantité produite Q_0. Sur la figure 8, nous détaillons la manière dont il faut substituer de l'input 2 par de l'input 1 pour aboutir à une même production Q_0 à partir d'une autre combinaison des inputs, désignée par *B*.

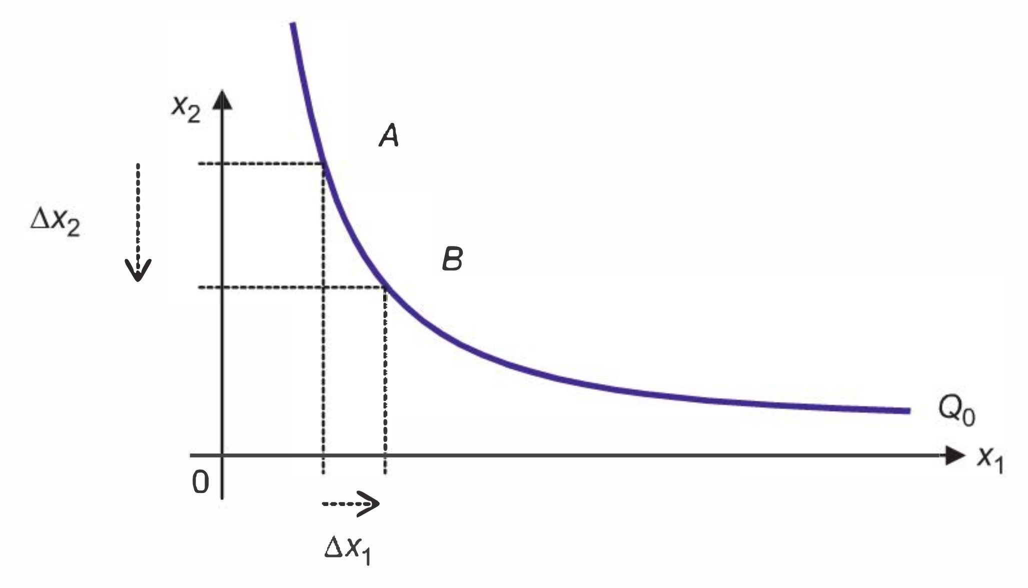


Figure 8 Substitution d'un input par l'autre ne modifiant pas la quantité produite d'output

Les proportions dans lesquelles il est possible de substituer de l'input 2 par de l'input 1 sans que cela ne modifie la quantité produite correspondent précisément au taux de substitution technique entre les inputs 1 et 2 :

$TST_{1/2} = \frac{\varnothing x_2}{\varnothing x_1}$. Ce ratio est négatif puisque soit le numérateur est négatif et le dénominateur positif (comme sur la figure où $\Delta x_2 < 0$ et $\Delta x_1 > 0$), soit le numérateur est positif et le dénominateur positif. Comme précédemment, nous voulons mesurer les proportions de la substitution en raisonnant, de préférence, sur des variations infinitésimales des quantités de facteurs combinées : ceci nous conduit à privilégier, non pas un taux de substitution technique mais un taux marginal de substitution technique, défini comme $TmST_{1/2} = \frac{dx_2}{dx_1}$. D'un point de vue graphique, le $TmST_{1/2}$ au point *A* est égal à la pente de la tangente à l'isoquant en ce point. C'est ce qui apparaît sur la figure 9.

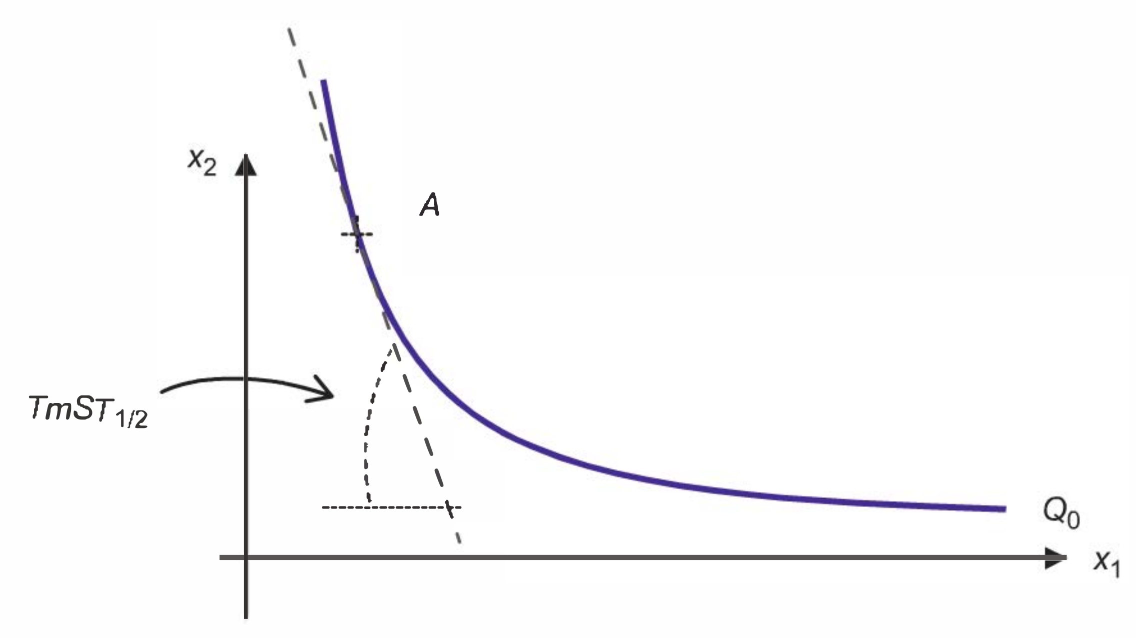


Figure 9 Visualisation du $TmST_{½}$

Mathématiquement, on établit que le *TmST* est égal, en valeur absolue, au rapport des productivités marginales.

Propriété

Le taux marginal de substitution technique entre deux inputs est égal (au signe près) au rapport des productivités marginales des inputs. Ainsi, $TmST_{½} = \frac{dx_2}{dx_1} = -\frac{Pm_1(x_1;x_2)}{Pm_2(x_1;x_2)}$

Nous avons caractérisé la manière dont une entreprise produit un output à partir de la combinaison de multiples inputs, implicitement introduits dans le processus de production en quantités variables. Nous avons négligé la question des infrastructures de production (bâtiments, machines, etc.) et plus généralement celle de tous les inputs mobilisés en quantités fixes. Cet aspect essentiel sera traité un peu plus loin. Nous allons tout d'abord nous interroger sur le choix de la meilleure – au sens de la moins coûteuse – combinaison d'inputs (implicitement utilisés en quantités variables). Pour cela, nous allons maintenant nous préoccuper des prix auxquels le producteur est susceptible d'acquérir les différentes unités d'inputs et calculer la dépense totale d'acquisition de ces inputs.

4 La combinaison optimale des facteurs de production : résolution du programme technique du producteur

Nous connaissons maintenant parfaitement la technologie de production. Nous avons compris que la production d'une certaine quantité d'output peut se faire non pas d'une, mais d'une infinité de façons. Alors, laquelle choisir ? Très naturellement, la réponse qui vient à l'esprit est que cela va dépendre du prix des différents inputs : si ceux-ci sont raisonnablement substituables, il sera opportun d'utiliser de façon très intensive celui (ou ceux) que l'entrepreneur pourra acquérir au meilleur prix et d'utiliser de manière peu intensive celui (ou ceux) que l'entrepreneur devra acquérir au prix fort. Jusqu'à présent dans ce chapitre, aucune référence à un quelconque prix n'était apparue : l'ensemble des considérations précédentes était de nature technique, au sens de l'organisation physique de la production.

4.1 Le prix des inputs

Il nous faut donc nous interroger sur la manière dont un entrepreneur peut se procurer des inputs utiles à sa production. Pour certains inputs, il lui suffit de passer commande à un fournisseur produisant un bien ou service très commun (prestations de nettoyages, location d'une flotte de véhicules, fourniture de consommables, etc.). Dans une telle configuration, le prix d'acquisition sera un « prix de marché » et les différentes entreprises clientes de ce fournisseur paieront toutes un prix identique. Pour d'autres inputs, il faut définir, avec le fournisseur, l'ensemble des caractéristiques du bien ou service, quasiment « fait sur mesure » pour le client (fabrication d'un moule, par exemple). Dans ce cas, le prix de l'input sera le fruit d'une âpre négociation et ne pourra pas être assimilé à un prix de marché car il sera issu d'une relation purement bilatérale.

D'autre part, il est concevable que le prix convenu soit dépendant des quantités achetées. Dans bien des cas, le prix est dégressif ; il peut aussi, plus rarement, varier positivement avec les quantités acquises : dans le cas de la fourniture d'une matière première rare, une forte demande exprimée par un client important peut peser significativement (à la hausse) sur le prix de la matière. On peut aussi rencontrer d'autres dispositifs de tarification non linéaires : le client peut se voir imposer par le fournisseur le paiement d'une partie forfaitaire avant de pouvoir acquérir, à un prix unitaire fixé, des unités d'input. C'est le cas, par exemple, des abonnements aux fournitures d'eau, d'électricité, de gaz, de téléphone et communications, etc. ou d'un distributeur qui paye une franchise à un fournisseur pour commercialiser ses produits.

Enfin, il faut mentionner que la capacité d'un fournisseur à fournir effectivement un client est entachée d'incertitude. À tout moment, la défaillance d'un fournisseur est un événement possible.

On l'a compris, la notion de prix unitaire lorsqu'il s'agit de l'acquisition d'un input n'est pas aussi simple à cerner que la notion de prix unitaire lorsqu'il est question d'un bien de consommation qu'une ménagère achète au supermarché. Par souci de simplicité, nous supposons néanmoins qu'il est possible de connaître avec précision le prix unitaire de chaque input, qu'à ce prix unitaire la firme peut acheter les inputs en quantité aussi importante qu'elle le souhaite et que ce prix unitaire demeure parfaitement indépendant des quantités acquises. Ainsi, dans le cadre de la production d'un output à partir de n inputs $x_1, x_2, ..., x_n$, les prix unitaires de ces inputs sont respectivement $w_1, w_2, ..., w_n$. Il s'agit maintenant d'établir quelle combinaison des différents inputs sera globalement la moins coûteuse pour produire un niveau d'output cible.

4.2 Le programme technique du producteur : la minimisation de la dépense d'acquisition des inputs

Selon une première lecture, ce niveau d'output cible est une quantité précise Q_0 que l'entrepreneur pense être en adéquation avec la demande émanant des consommateurs pour une période donnée. Selon une lecture plus

large, ce niveau de production cible est lui même une variable et le résultat du processus d'optimisation sera valide pour tout l'éventail des niveaux de production imaginables. C'est tout d'abord en conformité avec le premier niveau de lecture que nous allons raisonner. Le programme d'optimisation auquel est confronté le producteur est donc la minimisation de la dépense d'acquisition des facteurs de production sous contrainte d'atteindre un certain niveau d'output, en respectant les possibilités techniques résumées par la fonction de production. Le programme (dit « technique ») du producteur est donc :

$$\begin{array}{ll} \underset{x_1, x_2, \ldots, x_n}{Min} & w_1 x_1 + w_2 x_2 + \ldots + w_n x_n \\ s.c. & Q_0 \leq f(x_1 ; x_2 ; \ldots ; x_n) \end{array}$$

Comme nous l'avons dit plus haut, il n'y a ici, explicitement, que les inputs disponibles en quantité variable. Nous pourrions, dans ce programme d'optimisation, introduire la dépense relative aux inputs utilisés en quantité fixe. Ceci ne changerait rien au résultat quant aux quantités demandées d'inputs « variables » par le producteur.

Le programme ici posé est formellement un programme de minimisation sous contrainte que nous allons d'abord résoudre graphiquement dans le cas où seulement deux inputs sont combinés dans la fonction de production. La contrainte, que l'on va supposer être saturée à l'optimum, est alors $Q_0 = f(x_1 ; x_2)$, matérialisée par l'isoquant Q_0. La fonction objectif, la dépense d'acquisition des inputs, s'écrit quant à elle : $w_1x_1 + w_2x_2$. Nous allons représenter graphiquement cette dépense, que l'on peut désigner par D, dans le repère $(0 ; x_1 ; x_2)$. Réécrivons l'équation $D = w_1x_1 + w_2x_2$ comme $x_2 = \frac{D}{w_2} - \frac{w_1}{w_2} x_1$. Ceci est l'équation d'une droite (décroissante). Sur une telle droite apparaissent toutes les combinaisons des inputs 1 et 2 pour lesquels la dépense à consentir, afin de les acquérir, est D. On peut donc parler de droite d'iso-dépense ou droite d'iso-coût. Si l'on fait varier D, on fait monter ou descendre cette droite. Plus la droite est basse, plus la dépense D est faible. Sur la figure 10, on fait apparaître la droite d'iso-coût correspondant au niveau de dépense D. Son ordonnée à l'origine est $\frac{D}{w_2}$ et sa pente est $-\frac{w_1}{w_2}$.

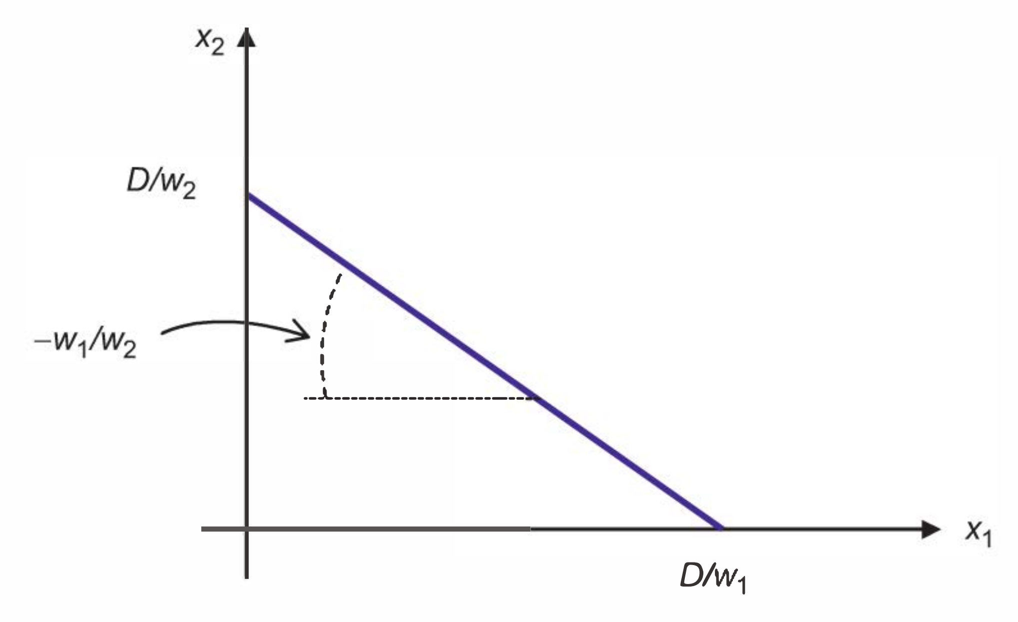


Figure 10 Droite d'iso-coût relative à une dépense d'acquisition des inputs D

Rechercher la solution du programme « technique » du producteur consiste donc à rechercher la plus basse droite d'iso-coût compatible avec l'isoquant Q_0. Sur la figure 11, on détermine l'optimum (x_1*, x_2*) en identifiant la plus basse droite d'iso-coût compatible avec l'isoquant Q_0.

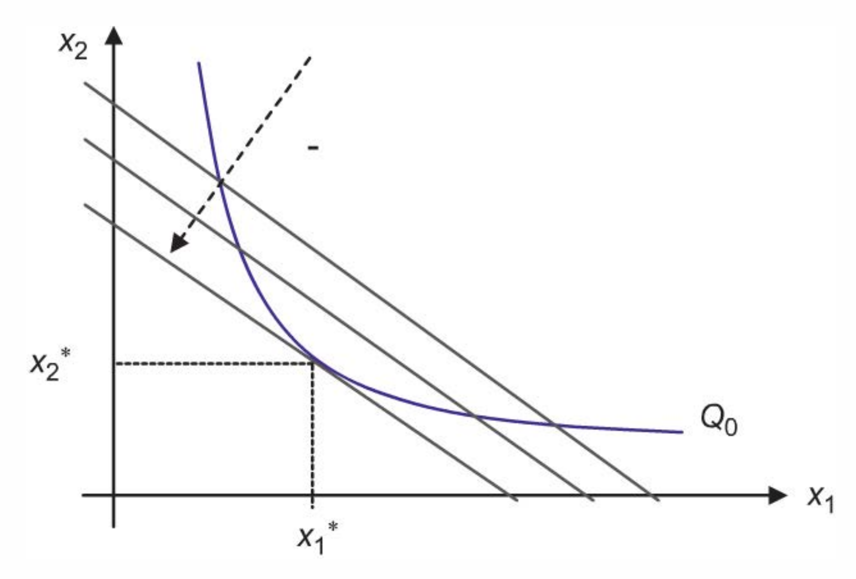


Figure 11 Optimum « technique » du producteur

Qu'observons-nous pour l'optimum (x_1*, x_2*) ? La plus basse droite d'iso-coût compatible avec l'isoquant Q_0 est précisément celle qui est confondue avec la tangente à l'isoquant considéré au point (x_1*, x_2*). Nécessairement donc, la pente de la tangente à l'isoquant (égale au $TmST_{1/2}$) est identique à

la pente de la droite d'iso-coût $-\frac{w_1}{w_2}$. Ainsi, ce qui caractérise la décision optimale de la firme est l'égalité :

$$TmST_{1/2} = -\frac{w_1}{w_2}$$

Caractérisation de la décision optimale du producteur (1)

La décision optimale du producteur est une combinaison de facteurs $(x_1^* ; x_2^*)$ telle que le taux marginal de substitution technique entre les inputs soit égal au rapport de leurs prix, encore appelé *prix relatif* (du facteur 1 relativement au facteur 2).

Comment interpréter cette propriété ? La décision optimale est atteinte quand la firme sera parvenue à faire coïncider le ratio qui mesure les proportions dans lesquelles il est possible de substituer un input par un autre avec le ratio des prix (qui exprime, en quelque sorte, les proportions dans lesquelles les conditions de commercialisation des facteurs « autorisent » la substitution d'un input par un autre). Il s'agit d'égaliser un ratio de substitution « technique » (le $TmST$) et un ratio de substitution « économique » (le rapport des prix).

En rappelant que $TmTS_{1/2} = -\frac{Pm_1(x_1 ; x_2)}{Pm_2(x_1 ; x_2)}$, nous pouvons aussi écrire que, à l'optimum :

$$\frac{Pm_1(x_1 ; x_2)}{Pm_2(x_1 ; x_2)} = \frac{w_1}{w_2}$$

ou, ce qui est équivalent :

$$\frac{Pm_1(x_1 ; x_2)}{w_1} = \frac{Pm_2(x_1 ; x_2)}{w_2}$$

Cette condition d'optimalité repose sur les termes $\frac{Pm_i}{w_i}$ $(i = 1, 2, \ldots, n)$ que l'on peut interpréter comme mesurant chacun le « surcroît de production,



par euro dépensé, obtenu grâce à l'unité supplémentaire de facteur i utilisé ». On peut, elle aussi, la dénommer condition « équi-marginale ».

Caractérisation de la décision optimale du producteur (2)

La décision optimale du producteur est une combinaison de facteurs (x_1* ; x_2*) telle que le surcroît de production, par euro dépensé, obtenu grâce à la dernière unité de facteur 1 utilisé soit égal au surcroît de production, par euro dépensé, obtenu grâce à la dernière unité de facteur 2 utilisé.

Si cette condition n'était pas vérifiée, le producteur réorganiserait la manière dont il combine les inputs et parviendrait à faire diminuer la dépense liée à leur acquisition. Par exemple, si le la dernière unité utilisée de l'input 1 engendrait un surcroît de production (par euro dépensé) supérieur à celui engendré par la dernière unité utilisée de l'input 2, le producteur accentuerait le recours au facteur 1 et réduirait le recours au facteur 2 dans la combinaison des inputs. Cet ajustement continuerait tant que les productivités marginales (par euro dépensé) ne seraient pas parfaitement identiques.

Pour aller plus loin

Résolution du programme technique du producteur

Le problème que nous avons résolu graphiquement ci-dessus peut être résolu de manière formelle. Rappelons, que, dans le cas où $n = 2$, le programme visant à minimiser la dépense d'acquisition des inputs sous contrainte d'atteindre un niveau de production Q_0 s'écrit :

$$\underset{x_1, x_2}{Min} \quad w_1 x_1 + w_2 x_2$$
$$s.c. \quad Q_0 \leq f(x_1 ; x_2)$$

Le programme ici posé est formellement un programme de minimisation sous contrainte dont la résolution nécessite des outils spécifiques que nous n'évoquons pas dans cet ouvrage. Au prix de quelques conditions supplémentaires, on peut affirmer qu'à l'optimum $Q_0 = f(x_1 ; x_2)$, ou, en d'autres

termes, que la contrainte est « saturée » (ce qui signifie que, pour minimiser sa dépense, le producteur se place rationnellement sur sa fonction de production et non en dessous). Au prix de quelques manipulations, on peut réécrire cette contrainte saturée sous la forme $x_2 = g(x_1 ; Q_0)$. Le programme devient :

$$\begin{array}{ll} \underset{x_1,x_2}{Min} & w_1x_1 + w_2x_2 \\ s.c. & x_2 = g(x_1;Q_0) \end{array}$$ qui peut être résolu en : $$\underset{x_1}{Min} \quad w_1x_1 + g(x_1;Q_0)$$

On obtient alors le couple de demande optimale de facteurs $(x_1^*; x_2^*)$ qui vérifie :

$$TMST_{1/2} = -\frac{\frac{\partial f}{\partial x_1}}{\frac{\partial f}{\partial x_2}} = -\frac{w_1}{w_2}$$

Si nous résolvons le programme relatif à la minimisation de la dépense d'acquisition des inputs combinés dans le cadre d'une fonction de production à n facteurs, c'est-à-dire $$\begin{array}{ll} \underset{x_1,x_2,\ldots,x_n}{Min} & w_1x_1 + w_2x_2 + \ldots + w_nx_n \\ s.c. & Q_0 \leq f(x_1; x_2; \ldots +; x_n) \end{array}$$, nous obtenons n demandes optimales de facteurs $(x_1^*, x_2^*, \ldots, x_n^*)$. La demande optimale de chaque facteur est une fonction des n prix et de la cible de production Q_0. On notera $x_i^*(w_1 ; w_2 ; \ldots ; w_n ; Q_0)$ la demande optimale d'input i. Il est alors possible de calculer précisément, la dépense (minimale) d'acquisition des facteurs (implicitement variables) relative à une production d'output Q_0 :

$$D^* = w_1x_1^*(w_1 ; w_2 ; \ldots ; w_n ; Q_0) + \ldots + w_nx_n^*(w_1 ; w_2 ; \ldots ; w_n ; Q_0)$$

4.3 Le coût fixe

La dépense consentie ci-dessus omet les éléments fixes du coût de production. Or, il ne serait pas raisonnable de prétendre que, pour une période de production donnée, tous les déterminants du coût sont des éléments variables. En effet, dans la quasi-totalité des processus, la production

nécessite l'utilisation de facteurs présents en quantité fixe, quel que soit le nombre d'unités finalement produites.

Prenons l'exemple de la production de véhicules automobiles : que l'on produise dans l'année 1 000 exemplaires ou 100 000 exemplaires d'un certain modèle, il est nécessaire de disposer d'une ligne de production complète (et il en est de même pour tous les fournisseurs de composants spécifiques au modèle considéré). Le coût correspondant à la « mise en place » de la ligne de production est un coût fixe, imputable à l'acquisition d'éléments de capital fixe. Souvent, les éléments du capital fixe engendrent la plus grande partie du coût fixe mais encore faut-il appréhender correctement ce qui est imputable, par période, à tel ou tel de ces éléments fixes sur lesquels s'appuie le processus de production. Les machines ont une durée de vie s'étalant sur plusieurs périodes d'analyse. Il faut, outre le coût de leur acquisition (principal et intérêts), comptabiliser le coût de leur entretien et de leur maintenance, le coût des modifications qu'elles auront à subir et retrancher, le cas échéant, le produit de leur revente. Ce coût complet est ensuite à répartir sur les différentes années d'exploitation. Ainsi, à chaque période, c'est un quantième seulement du coût complet des équipements fixes qu'il convient d'intégrer.

Dans certains secteurs, l'essentiel du coût fixe est le coût associé à la recherche & développement (R&D). Par exemple, dans l'industrie pharmaceutique, le développement d'une molécule satisfaisante nécessite de nombreuses années de recherche dont le coût total est considérable au regard de celui de la réalisation à grande échelle de cachets ou gélules la contenant. De même, les coûts de publicité et de marketing pèsent parfois d'un poids considérable au regard des éléments variables du coût (parfums, vêtements, etc.). Au titre de la R&D comme au titre du marketing, ces éléments de dépenses fixes sont largement supportés sous forme de dépenses d'acquisition de facteur « travail ». Il en va de même pour les rémunérations de l'encadrement et des services associés : quel que soit le volume d'activité, les salaires des responsables et des personnels des services industriel, commercial, administratif et financier, ressources humaines, import-export, approvisionnement, logistique, achats, etc. doivent être supportés.

4.4 Le coût total

Le coût fixe associé à la production d'une certaine quantité d'output Q_0 est donc un montant complexe à déterminer. Ce montant est sans lien avec la dépense d'acquisition des inputs variables (il peut être inférieur comme supérieur). Nous noterons désormais ce coût fixe *F*. Connaissant ce montant, il est alors possible de calculer la somme de la dépense d'acquisition des facteurs variables et du coût fixe relatifs à une production d'output Q_0. Cette somme, que nous appellerons « coût total de production » est égale à :

$$D^* + F = w_1 x_1^*(w_1\,;w_2\,;\ldots\,;w_n\,;Q_0) + \ldots + w_n x_n^*(w_1\,;w_2\,;\ldots\,;w_n\,;Q_0) + F$$

Il convient de noter que ce coût total de production est un coût « minimum » de production de la quantité d'output Q_0.

Nous sommes maintenant parvenus à nos fins dans l'optique du premier niveau de lecture : établir quelle combinaison des différents inputs sera globalement la moins coûteuse pour produire un niveau précis d'output cible Q_0. Comme nous l'avons dit, à un niveau de lecture plus large, ce niveau de production cible est lui-même une variable et le résultat du processus d'optimisation sera valide pour tout l'éventail des niveaux de production envisageables. C'est ce que nous allons matérialiser en caractérisant la fonction de coût total de production.

5 Fonction de coût total de production

Tout le raisonnement mené dans la section 4 peut être interprété comme une démarche permettant d'optimiser la combinaison des facteurs de production, non pas pour une seule et unique cible de production Q_0, mais pour toute quantité d'output produite *Q*. La première partie de l'expression du coût total de production D^* est dès lors assimilable à une fonction numérique qui indique, pour toute quantité *Q*, la valeur minimale de la dépense à consentir au titre de la part variable du coût de production. Cette partie est désormais désignée comme le coût variable de production.

On appelle **fonction de coût variable de production**, notée $CV(Q)$, la valeur minimale de la dépense à consentir pour acquérir les inputs combinés en quantités variable dans la production d'une quantité Q de l'output.

$$CV(Q) = w_1 x_1^*(w_1\,;w_2\,;\ldots;w_n\,;Q) + \ldots + w_n x_n^*(w_1\,;w_2\,;\ldots;w_n\,;Q)$$

Formellement, cette fonction dépend de $n + 1$ variables : les prix des n inputs w_1, w_2, ..., w_n et le niveau de la quantité produite de l'output Q. Il serait donc naturel de désigner la fonction de coût variable comme une fonction $CV(w_1\,;w_2\,;\ldots;w_n\,;Q)$, mais, dans la plupart des applications qui seront étudiées ensuite, le raisonnement se fait *à prix des inputs donnés* (c'est-à-dire invariants). Dès lors, pour ne pas surcharger les notations, on se contente d'écrire $CV(Q)$.

Pour passer de la fonction de coût variable à la fonction de coût total, il suffit d'ajouter F, la part fixe du coût. Comme décrit ci-dessus, établir la part du coût des infrastructures imputable à une période particulière est ardu. Établir, de même, ce qui, dans la masse salariale, constitue la dépense qu'il faut consentir indépendamment des quantités produites, est une tâche complexe. Nous supposons que toutes ces questions sont convenablement traitées et résolues et sommes alors en mesure de définir la fonction de coût total de production comme la somme des parts variable et fixe du coût.

On appelle **fonction de coût total de production**, noté $C(Q)$, la somme du coût variable et du coût fixe.

$$C(Q) = CV(Q) + F$$
$$= w_1 x_1^*(w_1\,;w_2\,;\ldots;w_n\,;Q) + \ldots + w_n x_n^*(w_1\,;w_2\,;\ldots;w_n\,;Q) + F$$

Rappelons que toute fonction de coût total de production demeure une fonction de coût **minimum** de production puisque cette fonction mesure la plus faible dépense à consentir pour parvenir, étant donné les caractéristiques de la technologie utilisée, à la production d'une quantité Q d'output. Sur la figure 12, nous représentons graphiquement une telle fonction

de coût total. Pour en comprendre la forme, nous allons définir, dans la dernière section, les notions de coût moyen (CM) et de coût marginal (Cm). La connaissance précise de ces fonctions nous permettra de discuter plus utilement de la forme de la fonction de coût total.

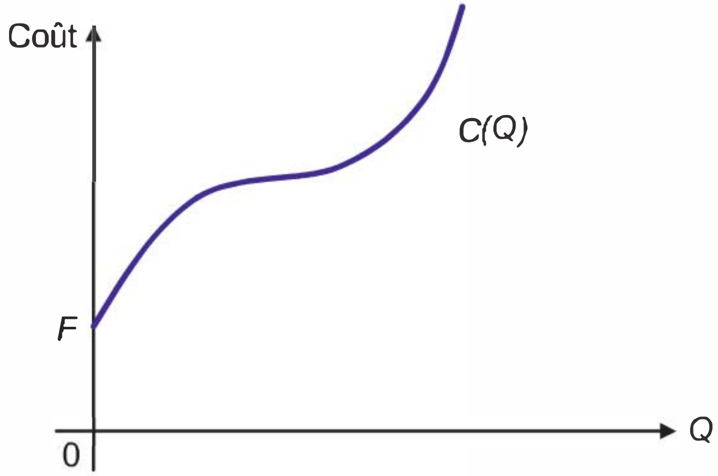


Figure 12 Fonction de coût total de production

6 Fonctions de coût moyen et de coût marginal

Connaissant la fonction de coût total de production d'une quantité Q d'un bien (ou service), nous souhaiterions exprimer le coût unitaire de fabrication de ce bien, ou en d'autres termes, le prix de revient de fabrication d'une unité ce bien. Ce prix de revient unitaire n'est pas constant. Il dépend de « l'échelle de production ».

6.1 Le coût moyen de production

Reprenons l'exemple de la production de véhicules automobiles sur une ligne de production. Si, pendant une période donnée, on produit un seul exemplaire du véhicule ou si l'on en produit 100 000, le prix de revient de fabrication du véhicule différera considérablement : il sera sans doute de plusieurs dizaines de millions d'euros dans le cas où l'on ne produit qu'un seul exemplaire, d'une dizaine de milliers d'euros si l'on en produit 100 000 unités. La fonction qui permet de connaître ce prix de revient de fabrication selon le nombre d'unités produites est la fonction de coût moyen de production.

On appelle **fonction de coût moyen ou coût unitaire de production**, noté $CM(Q)$, la fonction qui mesure, pour toute quantité produite, le prix de revient d'une unité produite.

$$CM(Q) = \frac{C(Q)}{Q}$$

Ce coût unitaire de production dépend fondamentalement de deux éléments :

- l'importance du coût fixe d'une part ;
- le surcroît de coût induit par la production de chaque unité supplémentaire d'autre part.

6.2 Le coût marginal de production

Si le coût fixe est connu une fois pour toutes, le concept de surcroît de coût induit par la production d'une unité supplémentaire est plus difficile à cerner. Il est surtout très différent selon la nature des biens produits et des technologies en jeu. Dans certains cas, ce surcroît de coût est constant : dans le cas de la production à la chaîne d'un équipement de petit électroménager, il coûte toujours la même chose de produire une unité supplémentaire, que l'on ait déjà produit 100 appareils ou 100 000. Dans d'autres cas, le surcroît de coût est croissant.

Par exemple, un agriculteur doit, pour produire des quintaux de blé supplémentaires, mettre en culture des terres de moins en moins fertiles ou de moins en moins faciles à exploiter (celles qui sont les moins planes ou les plus rocailleuses ou celles dont l'accès est le plus difficile) : chaque quintal supplémentaire coûte de plus en plus cher à produire.

Dans d'autres cas enfin, le surcroît de coût lié à la production d'une unité supplémentaire est décroissant. Ce phénomène de décroissance du coût de production de l'unité supplémentaire s'observe essentiellement pour les premières unités produites : en effet, au fur et à mesure, la production s'organise mieux, se rationalise. Chaque collaborateur devient plus habile dans

l'accomplissement de ses tâches et sait de plus en plus facilement répondre aux difficultés qu'il rencontre. Si le surcroît de coût induit par la production d'une unité supplémentaire peut être décroissant pour les premières unités produites, on s'attend néanmoins à ce qu'il devienne ensuite constant puis, vraisemblablement, qu'il commence à croître à nouveau à partir d'un certain seuil de production, ne serait-ce qu'en raison des difficultés liées à une taille de plus en plus grande de l'implantation industrielle (temps des déplacements sur le site, contraintes de stockage, surcoût lié à l'extension foncière).

On peut donner corps au concept de surcroît de coût lié à la production d'une unité supplémentaire d'output grâce à la notion de coût marginal.

On appelle **fonction de coût marginal de production**, noté $Cm(Q)$, la fonction qui mesure le surcroît de coût lié à la production d'une unité supplémentaire d'output.

$$Cm(Q) = \frac{\partial C(Q)}{\partial Q}$$

Il est à noter qu'il serait plus rigoureux d'invoquer le surcroît de coût engendré par la production d'une *quantité infinitésimale supplémentaire* d'output, ce que mesure en effet la dérivée première de la fonction de coût total.

6.3 La relation entre coût moyen et coût marginal

Dans le cas de l'économie numérique, le surcroît de coût lié à la production d'une unité supplémentaire est souvent infime : par exemple, le coût de « production » d'une unité supplémentaire de film téléchargé (légalement) sur internet est presque nul (pas exactement nul car il faut un peu d'électricité pour alimenter le serveur et le système de ventilation qui lui permet de fonctionner). Mais, si l'on peut considérer comme nul le surcroît de coût lié à la production d'une unité supplémentaire d'output, il n'en demeure

pas moins que ce surcroît de coût est constant et non décroissant : c'est le coût unitaire qui est décroissant, car l'essentiel du coût de production est un coût fixe (le coût de la production du film), progressivement « amorti » au fur et à mesure de la diffusion de l'œuvre.

Dans l'économie matérielle traditionnelle, on considère, dans la tradition microéconomique, que le coût marginal est d'abord décroissant puis croissant (ce qui est discutable comme nous l'avons fait remarquer ci-dessus). C'est ce que nous choisissons de représenter sur la figure 13. Nous remarquons sur cette figure un aspect essentiel du lien entre le coût moyen et le coût marginal : le second passe par le minimum du premier.

Propriété

Le coût marginal est sécant avec le coût moyen pour le niveau d'output tel que la production se fasse au coût unitaire minimum.

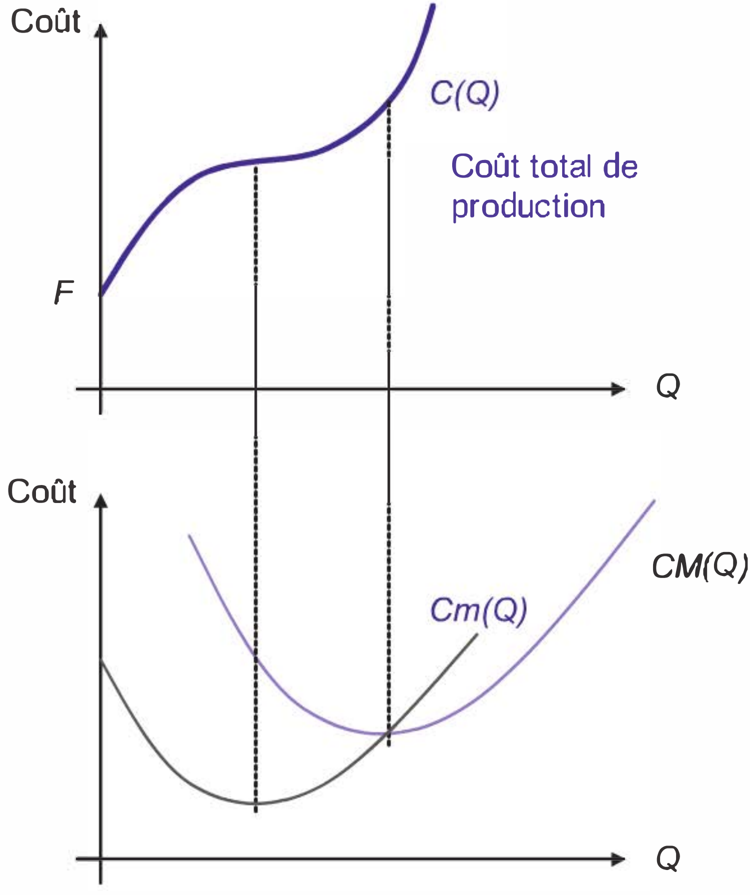


Figure 13 Fonctions de coût total, coût marginal et coût moyen

L'existence d'une partie d'abord décroissante pour le coût moyen n'est pas, en soi, dépendante de l'hypothèse d'un coût marginal initialement décroissant. Sur la figure 14, on constate que cette particularité s'observe aussi lorsque le coût marginal est constant.

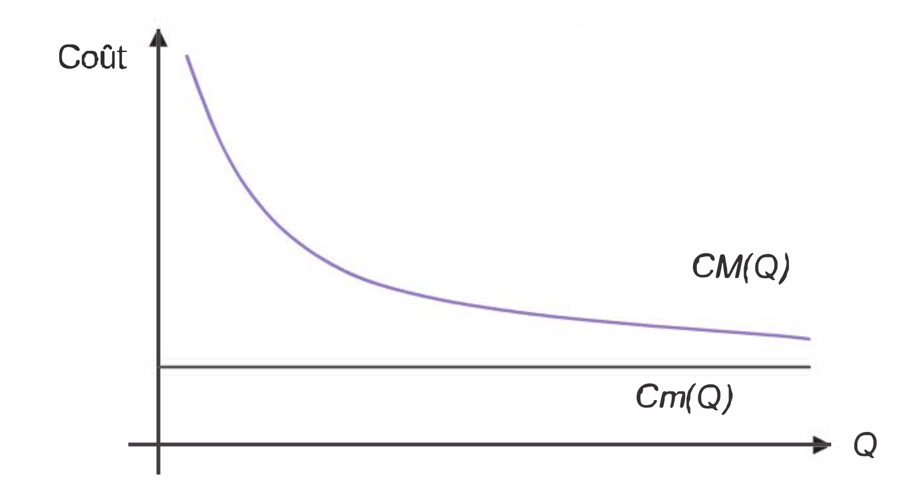


Figure 14 Coût marginal constant et coût moyen décroissant

Ce que l'on observe sur la figure 14 est lié à l'existence d'un coût fixe strictement positif. On parlera donc d'amortissement progressif du coût fixe. Le surcroît de coût lié à la production d'une unité supplémentaire est toujours le même, mais à chaque nouvelle unité produite, la charge du coût fixe est répartie sur un plus gros volume de production : c'est pourquoi le coût moyen est ici (toujours) décroissant et tend asymptotiquement vers le coût marginal. Ce qui est ici suggéré est plus proche de la réalité dans nombre de secteurs de production que ce qui apparaît sur le schéma de la figure 13. Concrètement, même si le coût marginal est faiblement croissant, l'importance des coûts fixes rend la fonction de coût unitaire « longtemps » décroissante (l'échelle de production à partir de laquelle le coût unitaire redevient croissant est parfois très supérieure aux débouchés effectifs de l'entreprise, en conséquence de quoi ce niveau de production n'est jamais atteint !). Ainsi, à contre-courant de la tradition des manuels de microéconomie, il convient de s'imaginer les courbes de coût unitaire des firmes comme fréquemment « principalement décroissantes » (voir figure 15).

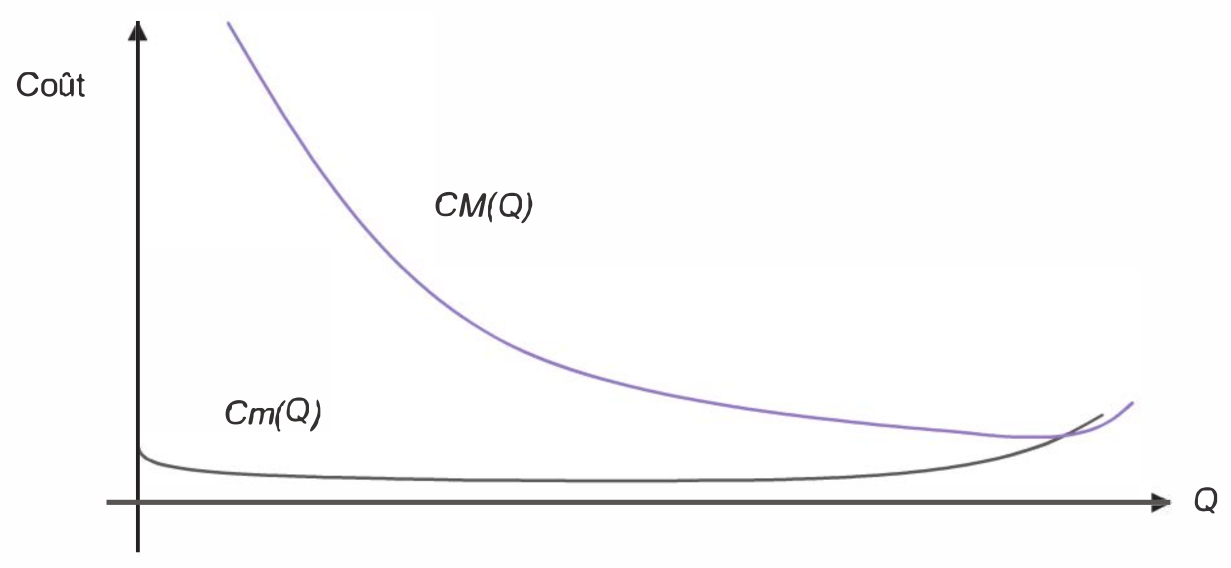


Figure 15

Ces propriétés des fonctions de coûts ont une importance décisive sur le degré de concentration du marché et l'intensité de la rivalité concurrentielle qui s'y exerce. Les remarques ci-dessus laissent entendre que le contexte de la concurrence « pure et parfaite », qui sera étudié dans le prochain chapitre, et qui repose sur des fonctions de coût du type de celles rencontrées sur la figure 13, n'est pas nécessairement le contexte le plus commun. Nous allons néanmoins l'étudier en détail car il constitue une sorte de point d'ancrage de la théorie économique, un cadre de référence universel.

7 L'offre de la firme en contexte de concurrence pure et parfaite

Mots-clés

Concurrence pure et parfaite, profit microéconomique, recette marginale, offre de court terme

Dans le précédent chapitre, nous avons vu comment la firme organise la production de son output au moindre coût. Quelle que soit la quantité qu'il s'avère opportun de produire (lors de la période de temps considérée), l'entreprise est en mesure d'établir comment il convient de réaliser cette production et à quel coût. Nous ne reviendrons pas sur la difficulté qu'il y a à établir, dans la vie réelle, les caractéristiques de cette relation entre quantité à produire et coût total (minimum) de production, et nous allons désormais nous tourner vers la manière dont l'output produit pourra être commercialisé auprès des consommateurs.

Dans le chapitre 4, nous avons précisé comment se caractérise la demande globale pour un bien ou service exprimée par l'ensemble des consommateurs présents sur le marché. Cette demande prend la forme d'une relation décroissante entre prix et quantité demandée. Il semblerait naturel de considérer que la firme a, quel que soit son domaine d'activité, une idée assez précise des caractéristiques de cette relation : un viticulteur qui exerce son activité depuis plusieurs années connaît approximativement le nombre d'hectolitres de vin qu'il peut espérer vendre en fonction du prix auquel il commercialise ses bouteilles. De même, un fabricant d'abris de jardin peut, lui aussi, estimer avec une certaine marge d'erreur le nombre d'abris de chaque type qu'il parviendra à vendre durant l'année selon les prix pratiqués.

Dans certains cas, la firme réalise qu'elle n'a guère le choix quant au prix qu'elle peut pratiquer : par exemple, un boulanger sait que s'il fixe à 2 € le prix unitaire de ses baguettes, il n'en vend pratiquement aucune. En revanche, s'il fixe à 0,50 € ce prix, il capte la clientèle de tout le quartier, et peu être même celle des quartiers voisins (à condition bien sûr que ses baguettes soient savoureuses). Mais à ce prix unitaire si bas, il ne parvient pas à couvrir ses charges. Ainsi il doit, peu ou prou, aligner son tarif sur celui pratiqué par les autres boulangers. Cette situation (où le prix unitaire d'un bien est à peu près identique, à qualité égale, d'un commerce à l'autre) ressemble à ce que les économistes appellent la « **concurrence pure et parfaite** », c'est-à-dire un contexte où aucune firme ne peut réellement pratiquer un prix différent de celui pratiqué par ses concurrents : afficher un prix supérieur ferait fuir tous les acheteurs tandis que, si, sur le principe, afficher un prix inférieur aurait la vertu de faire affluer toute la clientèle, il s'agirait, en réalité d'une pratique intenable car, à ce prix, la firme vendrait à perte.

L'émergence d'un tel contexte de concurrence pure et parfaite suppose l'existence de circonstances très particulières :

- Il faut que le bien ou service commercialisé soit quasiment identique d'un fabricant à l'autre, ce qui est assez rare (mais l'exemple de la baguette de pain est assez proche de cette intuition).
- Il faut que les firmes qui le produisent soient très nombreuses et de petite taille.

Imaginons qu'un seul et unique boulanger soit présent dans une ville et que le prochain boulanger soit à plusieurs dizaines de kilomètres. Le boulanger en place est tenté de pratiquer des prix élevés car sa clientèle est géographiquement captive. Peut-il escompter rester longtemps seul sur la place ? Dans un pays où la libre entreprise est la norme, il sera vite concurrencé par l'arrivée d'un ou plusieurs rivaux, attirés par la perspective de réaliser de bons gains. En vérité, il est même très peu probable de voir émerger une situation où un boulanger se retrouve en position de monopole dans une ville de bonne taille ; en effet, l'installation d'une boulangerie concurrente n'est pas en soi, un projet d'investissement de grande ampleur : il est assez simple et relativement peu coûteux de s'implanter. En résumé, dans le cas de la production et de la commercialisation de baguettes

de pain, il semble assez approprié de supposer que coexistent sur le marché une multitude d'entités de petites tailles qui fabriquent un bien quasiment identique, c'est-à-dire un ensemble numériquement conséquent de firmes évoluant dans un contexte de concurrence pure et parfaite.

Dans ce chapitre, nous allons établir comment une firme évoluant dans un tel environnement détermine sa décision de production. Dans une première section, nous allons d'abord préciser les hypothèses de la concurrence pure et parfaite.

1 Les hypothèses de la concurrence pure et parfaite

Nous avons, dans l'introduction ci-dessus, décrit un marché très particulier, caractérisé par la présence d'un très grand nombre de firmes de petite taille produisant toutes le même bien ou service. Un tel marché est considéré par les économistes comme « idéal » au sens où il rend possible de prévoir le comportement de chacun des acteurs de manière simple et, qu'à l'issue d'un certain délai, le fonctionnement du marché conduira à une allocation optimale (au sens de Pareto) des ressources. En réalité, l'existence d'un tel marché est très hypothétique. Les économistes le savent parfaitement. Mais en dépit de la faible vraisemblance d'un tel marché, ils s'efforcent d'en caractériser de manière très précise le fonctionnement supposé, de manière à établir le portrait d'un marché parfait, un **cadre de référence** pour toute l'analyse économique.

Un marché de concurrence pure et parfaite est un secteur de production d'un bien ou service pour lequel une multitude de firmes de très petite taille (**hypothèse d'atomicité**) produit un bien identique (**hypothèse d'homogénéité du bien produit**). Les firmes peuvent entrer et sortir librement de ce marché, sans coûts irrécupérables (**hypothèse de libre entrée**). Chaque firme est si petite qu'elle n'a aucune capacité à influer sur le prix auquel le bien ou service sera vendu sur le marché : le prix de marché s'impose à elle (**hypothèse de firme preneuse de prix, « price taker »**).

L'un des éléments les plus importants de ce descriptif est le fait que les firmes soient trop petites pour réellement peser sur la fixation du prix de marché. La question qui vient alors immédiatement à l'esprit est : comment, dès lors, se fixe ce prix de marché ? La réponse à cette question n'est pas immédiate et ne sera pleinement apportée que dans le chapitre 8. Nous supposerons dans un premier temps que ce prix unitaire p existe et que chacune des firmes présentes sur le marché peut connaître et appliquer ce prix. Si dans la ville, la tasse de café au comptoir vaut, de manière quasi généralisée, 1 euro et 50 cents, le nouveau cafetier sait, dès qu'il débute son activité, que c'est le prix qu'il « faut » pratiquer : un prix plus élevé serait excessif (les clients bouderaient son établissement), un prix plus faible ne lui permettrait pas de couvrir ses charges et de réaliser sa marge « normale ». Cette idée de marge « normale », très imprécise en l'état, est pourtant un paramètre essentiel pour délimiter ce qu'est le **profit** au sens microéconomique du terme.

2 La fonction de profit, fonction objectif du producteur

Ce que recherche le producteur est le profit. On peut nuancer cette affirmation en prétextant que certains entrepreneurs, avant de viser les bénéfices immédiats, cherchent d'abord à faire croître et prospérer leur entreprise. Il peut en effet y avoir un conflit entre la profitabilité immédiate et le souhait d'inscrire son entreprise dans la durée, en consolidant la structure par des investissements permanents, des innovations et une adaptation perpétuelle de l'activité. Mais ce que nous étudions ici a une portée limitée. Nous n'avons pas caractérisé l'entreprise et la technologie de production dans une optique temporelle. Tout se passe comme si le temps s'arrêtait (ou presque) à la fin de la période – mois, année – en cours. Nous n'envisageons donc l'objectif central de la firme que dans ce contexte simplifié. Ainsi, nous allons donc considérer que toute firme cherche à maximiser son profit microéconomique, défini comme la différence entre la recette totale, notée $RT(Q)$ et le coût total de production.

La fonction de profit, au sens microéconomique du terme, est la fonction qui mesure, pour toute quantité produite d'un bien ou service, la différence entre la recette totale tirée de la vente des unités de ce bien et le coût total de leur production.

Si l'on désigne par Q la quantité produite, on désignera la fonction de profit par :

$$\Pi(Q) = RT(Q) - C(Q)$$

Deux remarques importantes s'imposent à ce stade de nos développements :

- La microéconomie est une discipline qui s'inscrit dans la perspective d'une économie de « l'offre », c'est-à-dire dans la tradition intellectuelle résumée par Jean-Baptiste Say sous la forme d'une fameuse *loi des débouchés :* « Toute offre crée sa demande ». Ainsi, dès lors qu'une certaine quantité Q est produite, elle est réputée être vendue sans difficulté ! Ceci peut surprendre les lecteurs du XXI[e] siècle, nous qui constatons que les entreprises font parfois face à des stocks d'invendus considérables. Souvent même, c'est en raison de l'existence trop récurrente d'invendus que les entreprises périclitent… Mais rappelons que, dans le cadre hypothétique dans lequel nous nous trouvons, les firmes entrent et sortent du marché de manière très souple et que, si elles participent à la production du bien ou service, leur production demeure infime au regard des quantités échangées sur le marché. En résumé, dès lors que le bien homogène est proposé au prix unique du marché, il trouvera acheteur. Ceci explique que l'on puisse, sans précaution particulière, définir la fonction de profit à partir d'une seule variable Q : la quantité **produite** (dont le coût de production sera $C(Q)$) sera aussi la quantité **vendue** (qui engendre une recette totale $RT(Q)$). Cette approche, très facile à justifier dans le cadre de la concurrence pure et parfaite, sera aussi retenue dans le cadre de la concurrence imparfaite (monopole, chapitre 9 et oligopole, chapitre 10). L'hypothèse que la quantité produite par toute firme est aussi exactement la quantité vendue pourrait susciter, dans cet autre cadre, des interrogations et critiques beaucoup plus préoccupantes.

7

- La seconde remarque nous ramène à l'idée que, pour un entrepreneur, il existe une marge « normale » associée à la production des unités du bien ou service. Ceci signifie que le profit défini ci-dessus n'est pas un profit ou un bénéfice au sens usuel du terme. Il s'agit plutôt d'un surprofit, c'est-à-dire une rémunération au-delà de la rémunération normale de l'entrepreneur (ou du ou des propriétaire(s) de l'entreprise si le détenteur de l'entreprise n'est pas celui qui assure physiquement la mise en œuvre de la production et qui, dès lors, est *a minima*, bénéficiaire d'un salaire).

2.1 La notion de profit en microéconomie

Rappelons que la fonction de coût total de production inclut la rémunération de tous les inputs, variables ou fixes. En particulier, l'input capital doit être rémunéré. Au sein de cet ensemble figure l'ensemble des bâtiments, bureaux, ateliers et machines qui composent l'entreprise, ainsi que les stocks de biens intermédiaires prêts à être incorporés dans la production de nouvelles unités d'output. Ici vient d'être schématiquement décrit le socle sur lequel est défini le droit de propriété détenu par le ou les propriétaires de l'entreprise, ceux que l'on peut nommer, par simplicité, les actionnaires de la firme. Dans le coût total de production $C(Q)$ est donc déjà prise en compte la rémunération « normale » des actionnaires. Ainsi, si le profit microéconomique, défini comme la différence entre recette totale et coût total est strictement positif, cela signifie que les actionnaires percevront, en plus de leur rémunération « normale », une rémunération au-delà de la normale, une rémunération « anormale » en quelque sorte. Ce que nous décrivons ici est d'autant plus difficile à concevoir que l'on est coutumier des règles usuelles de la comptabilité : pour un comptable, le bénéfice est ce qui, tout en bas du compte de résultat, sera distribué aux détenteurs des actions, les propriétaires de l'entreprise. En microéconomie, le profit est une notion différente : certes, si un tel profit strictement positif existe, il sera à l'évidence distribué aux actionnaires, mais il leur sera distribué en plus du bénéfice « normal » qui leur revient. Le niveau « normal » du bénéfice qui revient aux propriétaires de l'entreprise dépend du secteur d'activité, du niveau des risques financiers encourus et donc du degré d'innovation qui a conduit à l'émergence du produit ou de son processus de production.

Il est important de conserver cette caractéristique présente à l'esprit ; nous verrons plus loin, en effet, que chaque marché de concurrence pure et parfaite est réputé conduire à une situation (dite d'équilibre de long terme) pour laquelle le profit de chaque firme sera nul. Ceci serait incompréhensible sans le décryptage présenté ci-dessus. En dépit de cette nullité annoncée du profit à long terme, l'objectif de chaque firme est néanmoins de le maximiser. C'est cette démarche que nous allons maintenant décrire dans la section suivante.

3 Résolution du programme « économique » du producteur

Puisque dans un cadre de concurrence pure et parfaite, l'entreprise ne possède aucune marge de manœuvre sur la fixation du prix de son produit, la seule variable sur laquelle porte sa décision est la quantité à produire Q (qui est supposée être immédiatement et intégralement achetée). L'objectif de la firme est donc de trouver la quantité d'output qui maximise son profit. Le programme économique du producteur s'écrit donc :

$$\underset{Q}{Max}\, \Pi(Q) = RT(Q) - C(Q)$$

Sous quelle condition la firme parvient-elle à maximiser le profit ? Sans entrer dans les détails mathématiques (aux termes desquels la dérivée première de cette fonction de profit doit être nulle), nous pouvons indiquer que le profit est maximisé lorsque la firme produit une quantité telle que la **recette marginale tirée de la vente de l'output est égale au coût marginal de sa production.** Pour comprendre cette condition, il faut d'abord préciser ce qu'est la recette marginale tirée de la vente de l'output.

La recette marginale tirée de la vente d'un bien ou service, notée $Rm(Q)$ est le surcroît de recette lié à la vente d'une unité supplémentaire de ce bien ou service.

$$Rm(Q) = \frac{\partial RT(Q)}{\partial Q}$$

7

Il est à noter qu'il serait plus rigoureux d'invoquer le surcroît de recette engendré par la vente d'une *quantité infinitésimale supplémentaire* d'output, ce que mesure en effet la dérivée première de la fonction de recette totale.

La quantité Q qui maximise le profit est donc celle pour laquelle $Rm(Q) = Cm(Q)$. Comme souvent en microéconomie, la condition d'optimalité prend la forme d'une relation indiquant que deux grandeurs marginales doivent être identiques. Cela signifie donc qu'*a contrario*, si les deux quantités en question ne sont pas égales, l'optimum ne sera pas atteint. C'est dans cette optique que nous nous proposons de décrypter la condition d'optimalité.

- Si la recette marginale était supérieure au coût marginal ($Rm(Q) > Cm(Q)$), cela signifierait qu'au regard de la quantité Q effectivement produite, produire une unité de plus rapporterait plus qu'il ne coûterait. Il faudrait donc produire plus. Il ne peut donc s'agir de l'optimum.
- Si la recette marginale était inférieure au coût marginal ($Rm(Q) < Cm(Q)$), cela signifierait qu'au regard de la quantité Q effectivement produite, produire une unité de plus coûterait plus qu'il ne rapporterait. Plus précisément, pour la quantité Q effectivement produite, la quantité seuil d'output au regard de laquelle un supplément de production rapporte plus qu'il ne coûte a été dépassée. Il faudrait donc produire moins. Il ne peut donc s'agir de l'optimum.

Seule la quantité d'output pour laquelle produire plus rapporte autant qu'il en coûte caractérise l'optimum.

Cette caractérisation de la décision optimale d'une firme est valable au-delà du contexte de la concurrence pure et parfaite (comme nous le verrons dans le chapitre 9). Dans le cadre particulier de la concurrence pure et parfaite, la recette marginale est précisément égale au prix p du bien ou service produit. En effet, si la firme n'a aucune capacité à peser sur le prix unitaire auquel le bien est vendu, elle a, en contrepartie, la « garantie » que chaque unité produite est vendue au prix unitaire p. Et si la firme augmente sa production d'une unité, cette unité supplémentaire est, elle aussi, vendue au prix p. Or ceci est précisément la définition de la recette marginale. Ainsi dans le contexte de la concurrence pure et parfaite, la condition d'optimalité se décline donc comme : $p = Cm(Q)$.

Pour aller plus loin

Résolution du programme « économique » du producteur

Puisque la firme est preneuse de prix ou « price taker », le prix unitaire p doit être considéré comme une variable exogène dans le calcul de la dérivée première de la fonction de recette totale. Ainsi, en concurrence pure et parfaite, la recette totale de la firme est :

$$RT(Q) = p \times Q$$

On peut alors écrire la condition de premier ordre conduisant à la solution du problème de maximisation du profit $\underset{Q}{Max}\, \Pi(Q) = RT(Q) - C(Q)$:

$$\frac{\partial \Pi}{\partial Q} = 0 \Leftrightarrow \frac{\partial RT(Q)}{\partial Q} - \frac{\partial C(Q)}{\partial Q} = 0 \Leftrightarrow Rm(Q) - Cm(Q) = 0 \Leftrightarrow Rm(Q) = Cm(Q)$$

$$\Leftrightarrow p = Cm(Q)$$

Cette condition d'égalité entre le prix et le coût marginal signifie que la firme doit choisir un niveau de production Q tel que le coût marginal de production soit, à cette échelle de production, précisément identique au prix qui s'impose à elle. Cette condition qui assure que le profit sera maximum ne garantit néanmoins aucunement que celui-ci soit positif. Mais une situation où une firme en concurrence pure et parfaite réaliserait un profit négatif (c'est-à-dire un bénéfice inférieur au niveau normal attendu dans ce secteur d'activité, peut être même négatif) n'est pas envisageable : la libre entrée et la libre sortie sans coût irrécupérable font qu'aucune firme ne sera présente sur le marché si elle n'obtient pas un profit positif ou nul.

Examinons donc les circonstances sous lesquelles le profit de la firme est positif ou nul. Il est possible d'établir que le profit est positif ou nul si et seulement si le prix est supérieur ou égal au coût unitaire de production, ou, en d'autres termes, si $p \geq CM(Q)$.

Représentons sur la figure 1 la relation entre prix et quantité offerte par la firme dans un repère avec la quantité en abscisse et le prix en ordonnée.

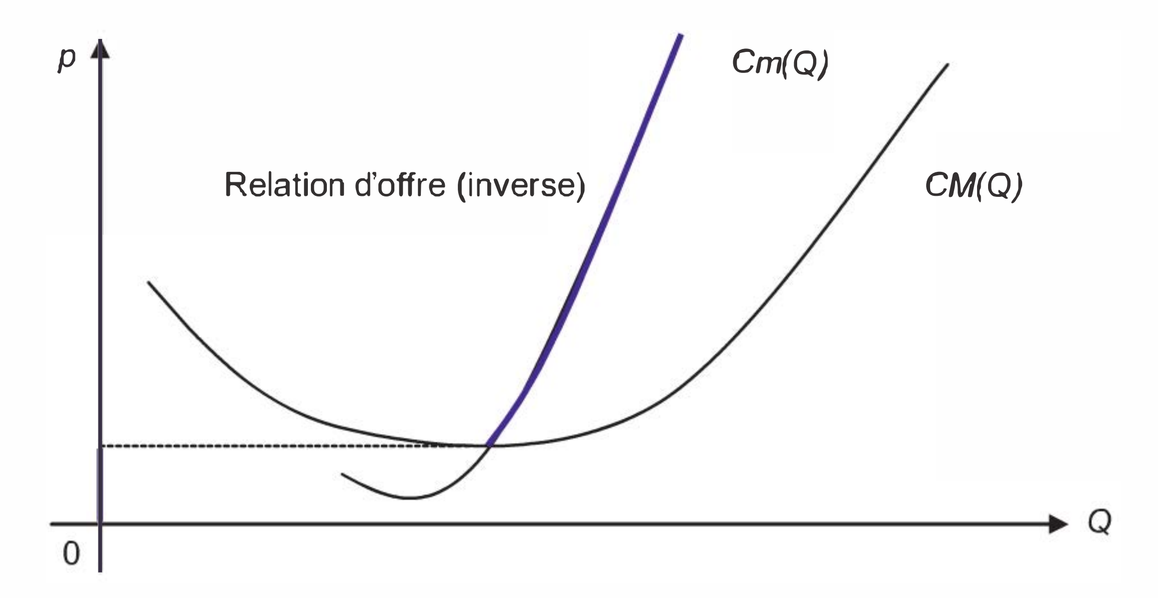


Figure 1 Relation entre prix et quantité offerte par la firme

Pour comprendre la figure 1, rappelons que l'on ignore, *a priori*, à quel niveau le prix de l'output se fixera. La courbe indique, pour chaque niveau possible de prix, quelle quantité la firme aura intérêt à produire et commercialiser. Bien entendu, si le prix unitaire ne permet pas de couvrir le coût unitaire de production, la firme décidera de ne rien produire. Or, rappelons qu'à l'intersection entre le coût marginal et le coût moyen se situe le niveau d'output pour lequel le coût unitaire est minimum : ainsi, la firme ne décidera de produire des unités d'output que si le prix unitaire de ce dernier excède le minimum de son coût moyen de production. C'est pourquoi, la relation entre prix et quantité offerte est **discontinue** : si le prix que le marché impose à la firme est strictement inférieur au minimum de son coût moyen, la quantité offerte par la firme est nulle ; si ce prix est supérieur ou égal au minimum de son coût moyen, la quantité offerte est positive, d'autant plus grande que le prix est grand, et « portée » par la courbe de coût marginal. Ce qui apparaît sur la figure 1 est ce que l'on appelle la relation d'offre « inverse » puisque c'est la quantité qui apparaît en abscisse. Il est aussi possible de faire apparaître la courbe d'offre, en inversant simplement les axes. C'est ce qui apparaît sur la figure 2.

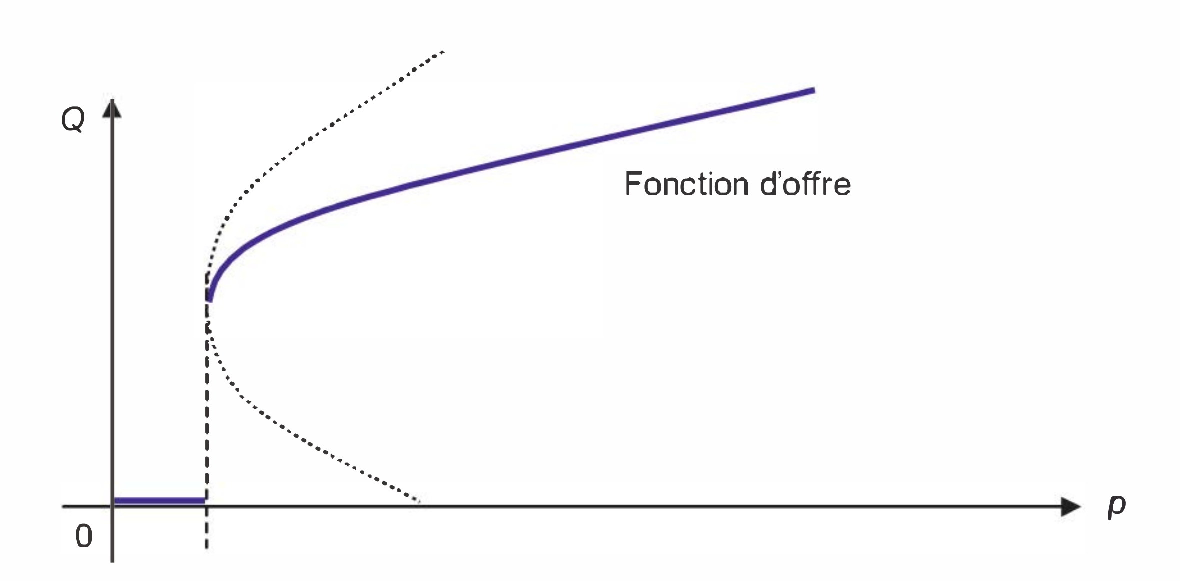


Figure 2 Fonction d'offre de la firme

Est-il possible que la firme produise alors même que son profit est négatif? Dans la tradition microéconomique, certains répondent par l'affirmative en distinguant un seuil de **rentabilité** et seuil de **fermeture**. Cette distinction n'est pas pertinente sous les hypothèses littérales de la concurrence pure et parfaite car elle n'est pas cohérente avec le postulat de libre entrée (et de libre sortie) – sans coût irrécupérable – des firmes. Néanmoins, dès que les caractéristiques du marché étudié s'éloignent un peu de ces hypothèses extrêmes, la distinction entre ces deux seuils a un sens.

3.1 Seuil de rentabilité et seuil de fermeture

Si un entrepreneur a engagé un investissement productif caractérisé par un coût fixe substantiel et seulement partiellement récupérable (ou récupérable seulement après quelques années d'exploitation) il est concevable que la firme puisse choisir de produire alors que son profit microéconomique est négatif. Prenons l'exemple d'un éleveur qui construit une étable sur son exploitation et qui l'équipe pour accueillir une centaine de têtes (de bovins, par exemple). Supposons en outre que la rentabilité de son exploitation ne soit assurée que s'il vend les carcasses aux bouchers à un prix moyen supérieur ou égal à 3 000 euros. Par malchance, le cours de la viande bovine fléchit et, après quelques mois d'exploitation, le prix

moyen d'une carcasse se stabilise à 2 500 euros. En dépit de ces circonstances qui empêchent l'éleveur de trouver une rentabilité à son exploitation, celui-ci ne peut pas sortir du marché immédiatement. Il faudrait qu'il trouve un repreneur (mais peu de repreneurs seront intéressés si le prix de la viande bovine est si bas), ou qu'il revende les installations. Les installations ne sont pas revendables comme on revendrait un véhicule ou un meuble : elles sont fixes et souvent attenantes au domicile de l'éleveur. Les démanteler est possible, mais ceci engendrerait un coût irrécupérable, car, même si quelques éléments meubles de l'équipement peuvent être revendus sur un marché d'occasion (pour un prix de revente plus faible que le prix du neuf) toute la dépense pour les éléments immeubles est perdue. Si l'on ajoute que l'éleveur ne peut pas, du jour au lendemain, cesser de nourrir ses bêtes, on comprend fort bien que l'hypothèse de libre sortie sans coût irrécupérable n'est pas vérifiée dans le cadre de cette activité. Alors, jusque quand continuer à produire ? La microéconomie répond en indiquant qu'il faut produire **tant que le prix permet de couvrir le coût variable moyen**. Rappelons (cf. chapitre 6) que le coût total de production est la somme d'un coût variable et d'un coût fixe ($C(Q) = CV(Q) + F$). Le coût variable moyen est le quotient du coût variable sur la quantité produite.

On appelle **fonction de coût variable moyen** de production, noté $CVM(Q)$, la fonction qui indique, pour toute quantité Q produite, quelle part, dans le prix de revient d'une unité produite, est imputable à l'acquisition des inputs combinés en quantités variable dans la production de l'output.

$$CVM(Q) = \frac{CV(Q)}{Q}$$

Cette fonction isole, dans le coût unitaire de production, ce qui est imputable aux « consommables ». Dans le cas de l'éleveur, il s'agit des dépenses de nourriture pour les animaux, d'électricité, d'eau, de médicaments et traitements sanitaires, etc. Il faut y ajouter tout ou partie du salaire que l'agriculteur se verse (même si nous avons dit, dans le chapitre précédent, que dans bien des cas une partie de la masse salariale des

entreprises s'apparentait à une charge fixe) et la rémunération qu'il verse à ses éventuels salariés. En revanche, les remboursements d'emprunts pour les bâtiments, les tracteurs ou les équipements de l'étable ne sont pas éléments du coût variable. Pourquoi (dans cette perspective, en marge des hypothèses de la concurrence pure et parfaite) peut-il être pertinent de continuer à produire si le prix de vente couvre le coût variable moyen ? Parce que, dans cette circonstance, la perte est moindre que si l'on abandonnait radicalement l'activité productive. En effet, en cessant du jour au lendemain l'activité productive, plus aucune recette ne serait perçue par l'exploitant alors que celui-ci doit toujours supporter la part fixe du coût (les remboursements d'emprunts). La perte subie serait supérieure à celle enregistrée en continuant à produire et vendre les bêtes si, répétons-le, le prix de vente n'est pas inférieur à ce qui, dans le coût unitaire, est imputable aux dépenses variables pour la nourriture des animaux, l'eau, l'électricité, les médicaments, etc. Bien sûr, si le prix unitaire ne couvre même plus cette part variable, il faut cesser la production immédiatement. On distingue donc un seuil de rentabilité et un seuil de fermeture :

- Tant que le prix de vente unitaire du bien couvre le coût moyen de production, le profit microéconomique est positif et l'on se situe au-dessus du seuil de rentabilité.
- Si le prix de vente unitaire du bien est en dessous du coût moyen mais au-dessus du coût variable moyen de production, la production continue mais le profit microéconomique est négatif. On se situe entre le seuil de rentabilité et le seuil de fermeture.
- Si le prix de vente unitaire du bien est en dessous du coût variable moyen, la production cesse et le profit microéconomique est négatif car l'entreprise continue à supporter la part fixe du coût. On se situe sous le seuil de fermeture.

Les seuils de rentabilité et de fermeture apparaissent sur la figure 3. La fonction d'offre inverse est la courbe en trait gras. On observe donc que la firme produit et propose à la vente des unités de son produit lorsque le prix est inférieur au minimum de son coût unitaire, tant qu'il demeure au-dessus du coût variable moyen.

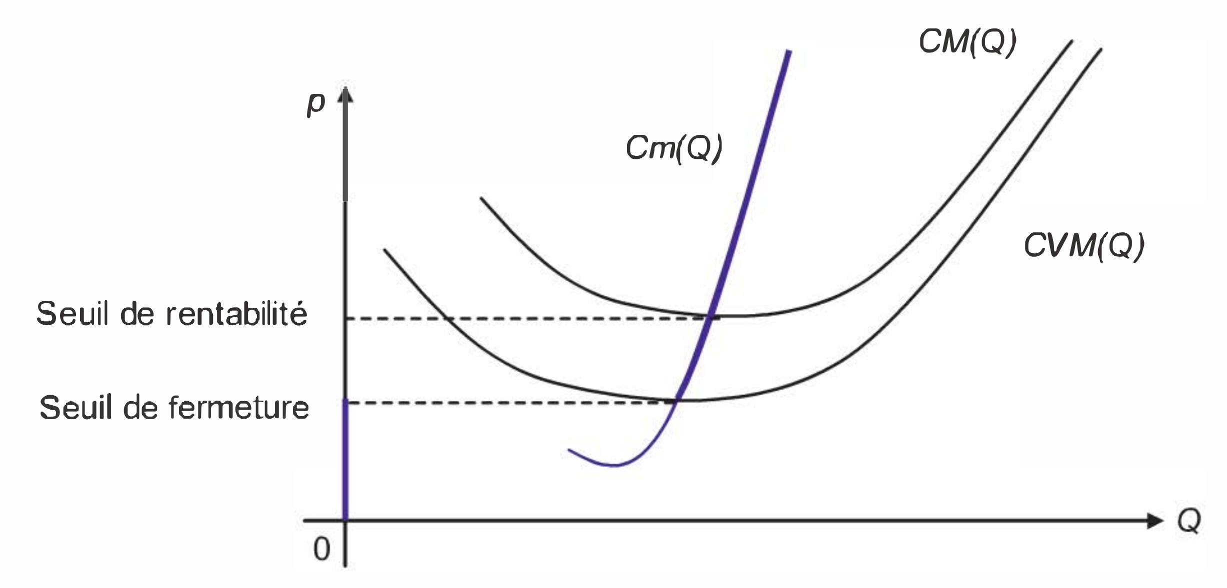


Figure 3 Seuil de rentabilité et seuil de fermeture

Pour aller plus loin

Détermination du seuil de fermeture

Si l'on note Π^p le profit réalisé par la firme en produisant et Π^{np} le profit réalisé en ne produisant pas, on peut comparer les « gains » (en réalité des pertes) :

$$\Pi^p \geq \Pi^{np}$$
$$\Leftrightarrow p \cdot Q - CV(Q) - CF \geq -CF$$
$$\Leftrightarrow p \cdot Q - CV(Q) \geq 0$$
$$\Leftrightarrow p \cdot Q \geq CV(Q)$$
$$\Leftrightarrow p \geq \frac{CV(Q)}{Q}$$
$$\Leftrightarrow p \geq CVM(Q)$$

Comme annoncé plus haut, la firme préfère continuer à produire (plutôt que cesser totalement sa production) tant que le prix reste supérieur au coût variable moyen.

Parfois, dans certaines périodes d'activité, il est donc opportun de vendre à perte car ces pertes seront, au total, moindres que si l'on ne produisait pas. Dans le secteur du tourisme, un tour opérateur, un établissement hôtelier ou une compagnie aérienne peuvent délibérément choisir de brader leurs prestations pour « remplir » séjours, hôtel ou avions, en ayant parfaitement

conscience des pertes qu'engendrera cette activité aux tarifs où seront commercialisées les prestations. Ces pertes sont néanmoins moindres que celles qu'engendrerait la non-production : les coûts d'exploitation d'un hôtel ne sont pas considérablement plus élevés lorsque l'hôtel est plein que lorsqu'il est presque vide ; il en va de même pour le coût de production d'un trajet aérien lorsqu'il n'y a qu'un seul passager au regard d'un vol complet. Enfin rappelons qu'entre seuil de rentabilité et seuil de fermeture c'est le profit microéconomique qui est négatif. Il est donc même possible que sous le seuil de « rentabilité » la firme réalise néanmoins quelques maigres bénéfices au sens comptable du terme.

4 L'agrégation des offres individuelles dites de court terme

Revenons aux strictes hypothèses de la concurrence pure et parfaite et oublions la notion de seuil de fermeture. Nous sommes donc confrontés à une firme qui offre des unités de bien ou service dès lors que le prix unitaire (que le marché lui impose) couvre son coût unitaire de production (c'est-à-dire un prix au-dessus du seuil de rentabilité). Cette firme n'est pas seule. Il y a en effet, dans ce secteur d'activité, une multitude de firmes qui produisent toutes exactement le même produit. Nous allons tout d'abord supposer que ces firmes sont susceptibles de produire avec des technologies différentes et donc avec des fonctions de coût différentes. Par exemple, certaines se sont installées depuis longtemps et utilisent une technologie ancienne. D'autres, en revanche, viennent de s'installer et ont opté pour un processus technologique plus récent. Au prix de marché (dont on ignore encore comment il se fixe), certaines des firmes pourront réaliser un profit microéconomique positif ou nul et d'autres ne participeront pas au marché car elles ne réussiraient pas à couvrir leur coût unitaire. Nous allons tout d'abord supposer que le nombre de firmes susceptibles de produire le bien est donné. Ceci correspond à la notion de **court terme** en microéconomie. Par opposition, le long terme désignera le délai à l'issue duquel le nombre de firmes se sera ajusté pour parvenir à l'équilibre de concurrence pure et parfaite. Cette notion d'équilibre sera examinée dans le chapitre 8, ainsi que le principe de la formation du prix de marché de l'output. Dans ce chapitre,

nous nous contentons de supposer qu'il existe un prix (de court terme) de l'output et qu'il existe un nombre donné de firmes susceptibles de le produire. Comment peut-on agréger les offres de toutes ces firmes et parvenir à une fonction d'offre globale ? La réponse est finalement assez intuitive : toutes les firmes qui, au prix de court terme, réalisent un profit microéconomique positif ou nul, offrent un certain nombre d'unités d'output. À chaque niveau de prix de court terme, il est possible d'établir le nombre de firmes prenant part au marché (c'est-à-dire produisant au moins une unité d'output) et la quantité totale offerte. La manière dont s'agrègent les offres individuelles de court terme est représentée sur la figure 4.

Sur la figure 4, on représente l'agrégation des offres individuelles d'un marché où seulement trois firmes sont susceptibles de produire de l'output. Ces trois firmes recourent à des technologies différentes et supportent des coûts différents. La plus « efficace » est la firme 3, dont les fonctions de coût sont les plus basses : pour un même niveau d'output à produire, c'est la firme 3 qui parviendra à produire ces unités au coût le plus faible. À l'opposé, c'est la firme 2 qui est la moins efficace, c'est-à-dire celle qui produit au coût le plus élevé. Supposons que le prix (de court terme) s'imposant aux firmes est p_0. Comme on le voit, seules les firmes 3 et 1 sont susceptibles, à ce prix, de produire en couvrant leur coût unitaire. *A contrario*, à ce prix p_0, la firme 2 ne participera pas à la production : elle sortira du marché (sans délai et sans coût irrécupérable). La quantité totale produite par ces deux firmes sera Q_0 précisément égal à la somme des quantités offertes par les firmes 3 et 1.

En reproduisant le même raisonnement pour toute valeur possible du prix de court terme, on peut construire la courbe d'offre globale de court terme $Q^s(p)$ (en inversant la relation fonctionnelle dont on obtient la représentation graphique sur la figure 4).

Remarquons qu'à court terme, le profit des firmes prenant part à la production de l'output peut être strictement positif. C'est ici ce que nous observons pour les firmes 3 et 1 lorsque le prix de court terme de l'output est p_0 : en effet, le coût unitaire de production des unités d'output qu'elles décident d'offrir est inférieur à ce prix p_0. Ainsi leurs profits sont strictement positifs. C'est ce qu'il nous est donné d'observer sur la figure 5.

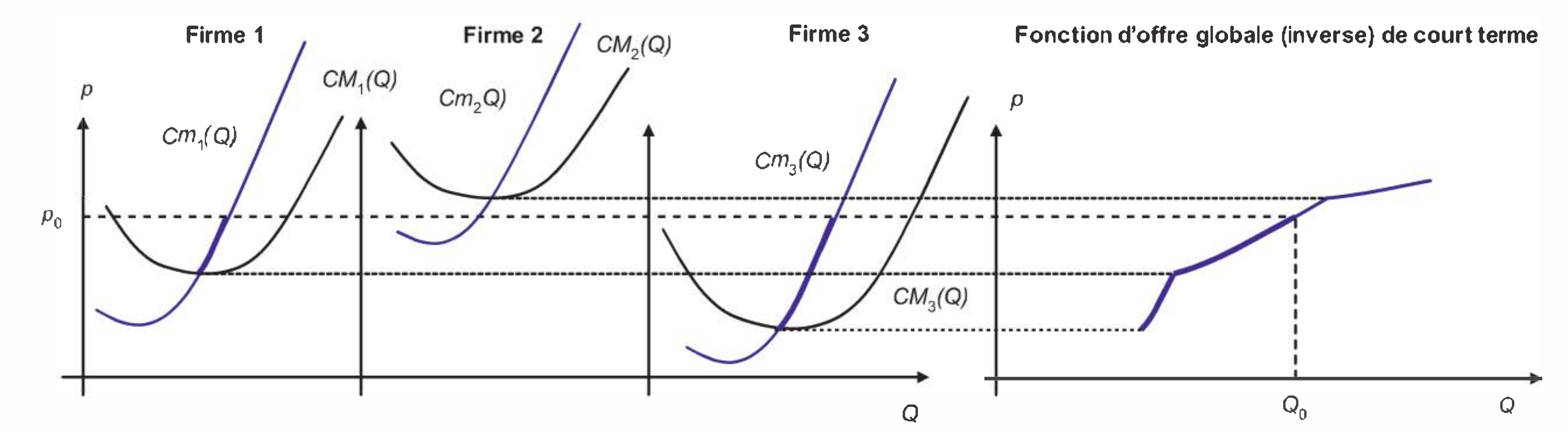


Figure 4 Agrégation des offres individuelles de court terme

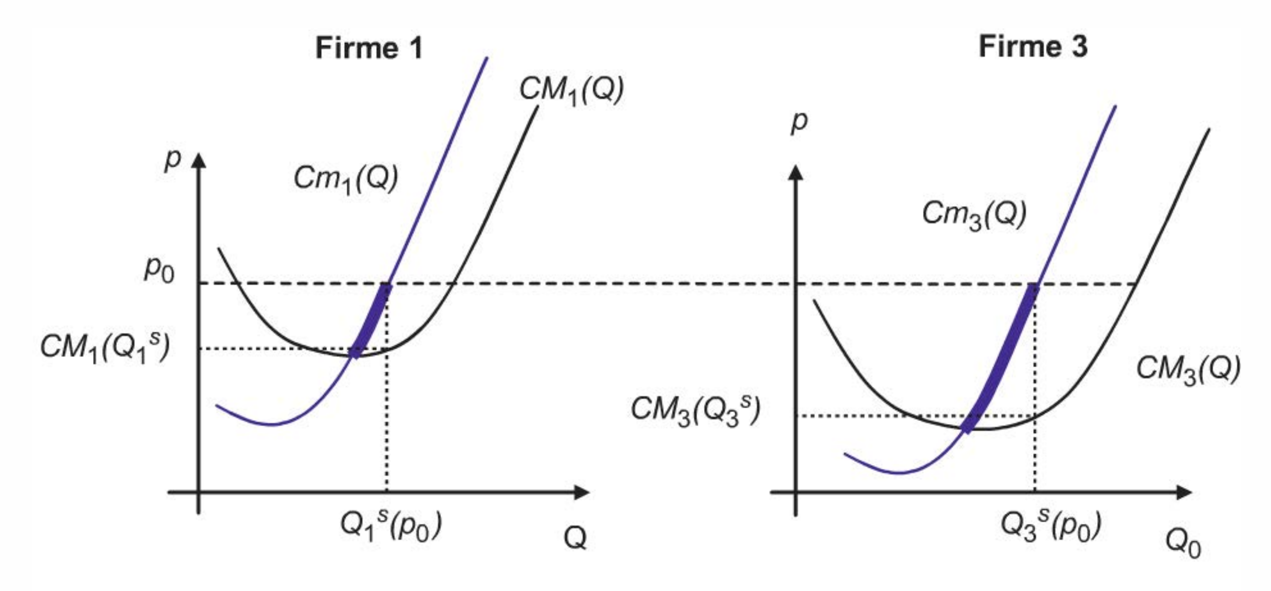


Figure 5 Les profits des firmes peuvent être strictement positifs à court terme

Nous connaissons maintenant l'offre globale de court terme sur ce marché de concurrence pure et parfaite. Il reste à comprendre ce qui se passe à long terme, c'est-à-dire à l'issue d'un délai où de nouveaux concurrents seront entrés sur le marché. En effet, la perspective de réaliser un profit microéconomique strictement positif va nécessairement attirer de nouvelles firmes. Ce que nous allons examiner dans le chapitre 8 est la manière dont ce mouvement va nous conduire à un prix d'équilibre dit de long terme. Selon l'intuition première, ce prix devrait dépendre de l'interaction entre l'offre et la demande. Or, en confrontant les caractéristiques de l'offre et de la demande, nous découvrirons que si le nombre de firmes présentes à long terme ne peut être déterminé qu'en fonction de l'intensité de la demande exprimée par les consommateurs, il est possible de connaître le prix d'équilibre de long terme du marché sans connaître le niveau de cette demande globale. Plus précisément, nous allons comprendre, qu'une hypothèse jusqu'ici passée sous silence – l'hypothèse de « diffusion technologique » – conditionne la cohérence globale de la construction théorique qu'est le modèle de concurrence pure et parfaite.

8 L'équilibre d'un marché

Mots-clés

Offre globale, court terme, long terme, équilibre du marché, modèle du Cobweb

Dans le chapitre 4, nous avons détaillé la manière dont se forme la fonction de demande globale d'un bien ou service exprimée par l'ensemble des consommateurs susceptibles de prendre part à la consommation du bien. Dans le chapitre 7, nous avons établi comment, dans une optique de court terme, les entreprises en concurrence pure et parfaite décident des quantités à produire du bien ou service. Il faut maintenant envisager l'interaction entre cette demande et cette offre, dans une optique de long terme. Chacun pensera spontanément que c'est dans le cadre de la confrontation entre l'offre et la demande que se détermine le prix du marché. C'est une louable intuition mais elle est partiellement incorrecte dans le contexte (très particulier) de la concurrence pure et parfaite. Si la quantité effectivement vendue par les producteurs aux consommateurs est, en effet, le résultat de la confrontation entre offre et demande, le prix d'équilibre est seulement déterminé par les conditions technologiques, c'est-à-dire les conditions dans lesquelles les firmes peuvent réaliser la production de l'output. Les hypothèses de la concurrence pure et parfaite conduisent à une efficacité productive absolue, qui garantit que le prix est exactement égal au coût unitaire minimum (indépendamment de l'intensité plus ou moins élevée de la demande). Dès lors, ce sont bien les caractéristiques de la technologie de production la plus performante (c'est-à-dire la moins coûteuse) qui déterminent le niveau auquel se fixe le prix de marché « définitif ». Mais pour que cette construction soit totalement cohérente, il faut que la tech-

nologie se diffuse « spontanément » auprès de la multitude des firmes présentes sur ce marché ; ainsi, chacune des firmes sera en mesure de produire similairement au coût unitaire minimum. Cette idée de diffusion technologique peut correspondre à l'élimination de firmes recourant à des technologies archaïques ou exagérément coûteuses, au profit de nouveaux arrivants adoptant la technologie la plus performante. Une fois encore, aux termes des hypothèses de la concurrence pure et parfaite, ce processus d'entrées et sorties de firmes se fait de manière très souple et sans coût irrécupérable.

Après avoir décrit le mécanisme conduisant à l'équilibre de long terme d'un marché de concurrence pure et parfaite, nous décrirons les caractéristiques de l'équilibre d'un marché quelconque, puis nous discuterons de la stabilité de l'équilibre : est-ce que, suite à un « choc », une perturbation, les forces du système économique ramène le prix au niveau initial (c'est-à-dire au prix d'équilibre) ?

1 L'équilibre d'un marché de concurrence pure et parfaite

Dans le chapitre 7, nous avons étudié la décision d'offre de court terme des firmes présentes sur un marché de concurrence pure et parfaite. Dans cette optique de court terme, les firmes différaient par leurs fonctions de coût. Pour un niveau donné du prix de l'output, certaines d'entre elles, trop peu performantes, ne produisaient pas car elles étaient incapables, à ce prix, de couvrir leur coût unitaire. D'autres, à l'inverse, réalisaient un profit microéconomique strictement positif, autant dire des gains alléchants. Que peut-il se passer à l'issue d'un certain délai ? D'autres entrepreneurs seront désireux de se positionner sur ce marché où, à condition de choisir la bonne technologie, les gains sont assurés. Sous les hypothèses de la concurrence pure et parfaite, il n'y a aucune raison qu'un nouvel entrant ne puisse pas se doter de la technologie la plus performante. En effet, aucun brevet, aucune technologie secrète, aucune restriction d'accès aux différents procédés et techniques ne sont réputés exister.

L'entrée de nouvelles firmes adoptant la technologie la plus efficace a une conséquence immédiate : le prix auquel l'output peut être vendu va baisser avec l'entrée des nouvelles firmes et donc le profit que celles-ci vont réaliser sera plus faible qu'escompté. Reprenons le graphique sur lequel était représentée la fonction d'offre globale (inverse) de court terme (c'est-à-dire pour un nombre initialement donné de firmes) et faisons désormais apparaître la relation entre prix et quantité globalement demandée du bien ou service (la demande inverse, cf. chapitre 4). Avec, comme dans l'illustration retenue dans le chapitre 7, une courbe d'offre issue des décisions d'un nombre très limité de firmes aux technologies hétérogènes, le prix susceptible de faire coïncider l'offre et la demande sera plutôt élevé, et ce prix constituera un signal attractif pour de nouveaux entrepreneurs : en utilisant la technologie la plus performante, tout nouvel entrant a la garantie de réaliser un profit microéconomique strictement positif. C'est que nous voyons apparaître sur la figure 1.

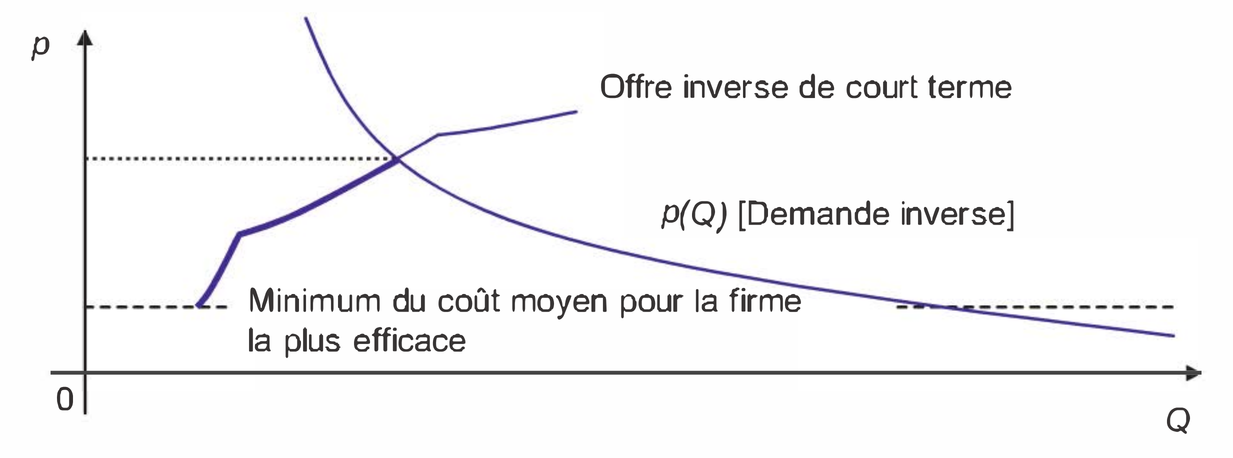


Figure 1 Offre de court terme et demande

Attirées par la perspective de réaliser un profit strictement positif, de nouvelles firmes affluent donc sur le marché. Ces entrées sur le marché vont considérablement modifier la forme de la fonction d'offre globale (inverse). En effet, tous les entrants vont adopter la technologie la plus performante (il serait absurde, pour un entrant, d'adopter une technologie moins bonne). Reprenons le raisonnement du chapitre 7 où l'offre globale se déterminait comme la somme des quantités offertes par

chacune des firmes susceptibles de produire en couvrant son coût unitaire de production, et considérons maintenant l'offre de nouveaux entrants qui se sont tous « alignés » sur le processus de production le plus efficace. Nous pouvons observer, sur la figure 2, que la fonction d'offre globale qui se forme possède d'abord une longue partie horizontale : chacune des firmes présentes offre une même quantité d'output dès que le prix atteint le seuil (désormais commun) que représente le minimum du coût moyen de production. Si le prix dépasse ce seuil, chacune des firmes augmentera elle-même son offre (et chacune offrira, pour chaque niveau de prix, une même quantité que ses concurrentes). Il résulte donc, à la suite de la partie horizontale, une lente croissance de la courbe d'offre (inverse).

Combien de firmes vont entrer sur ce marché ? De nouveaux entrants affluent tant qu'il est possible de réaliser un profit strictement positif. Ceci nous renseigne sur ce que sera le **prix de marché** à l'issue du mouvement d'entrées sur le marché : nécessairement, le flux d'entrée se tarira dès lors que le profit deviendra nul, c'est-à-dire lorsque **le prix aura atteint le plancher que constitue le minimum du coût moyen**. Le délai à l'issue duquel le profit devient nul est appelé le **long terme**. Quant au nombre N de firmes présentes sur le marché à long terme, il est déterminé par l'intersection de l'offre (inverse) et de la demande (inverse). C'est ce qui apparaît sur la figure 3.

Sur la figure 3, le nombre N de firmes présentes sur le marché à long terme est désormais connu. Au prix p_0 égal au minimum du coût moyen, chacune des N firmes offre une certaine quantité Q_0, ce qui conduit à une offre totale sur le marché $N \times Q_0$. Le nombre N est précisément le nombre tel que la demande exprimée pour le bien ou service produit, s'il est vendu au prix p_0, soit $N \times Q_0$. Ainsi, c'est bien l'interaction entre l'offre et la demande qui détermine la quantité totale échangée sur le marché (et donc le nombre d'entreprises présentes à long terme), mais, intrinsèquement, le prix d'équilibre de ce marché est déterminé par les conditions technologiques car précisément égal au minimum du coût unitaire de production pour une quelconque firme utilisant la technologie la plus efficace.

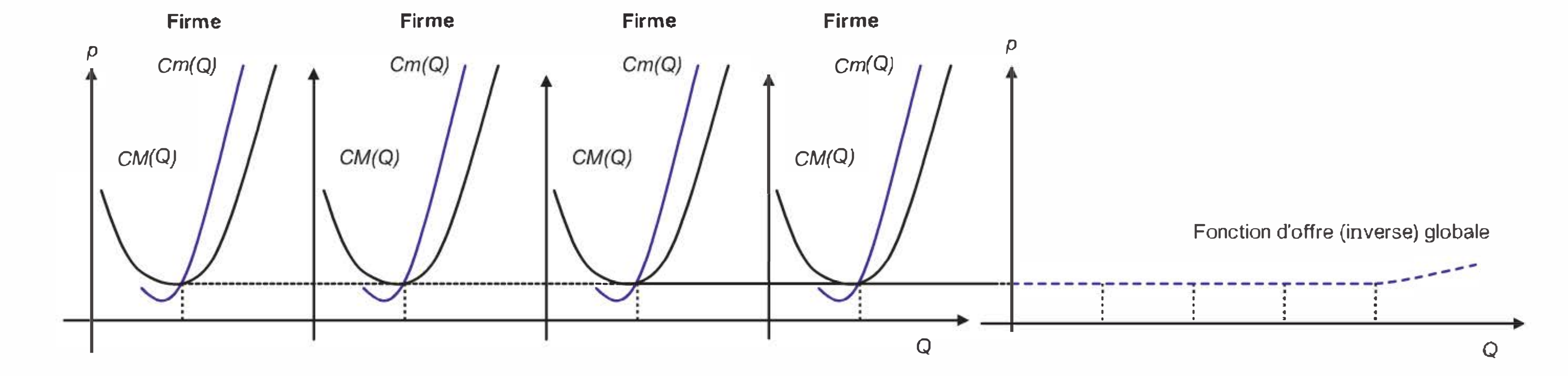


Figure 2 Agrégation des offres individuelles de firmes ayant toutes adopté la technologie la plus efficace

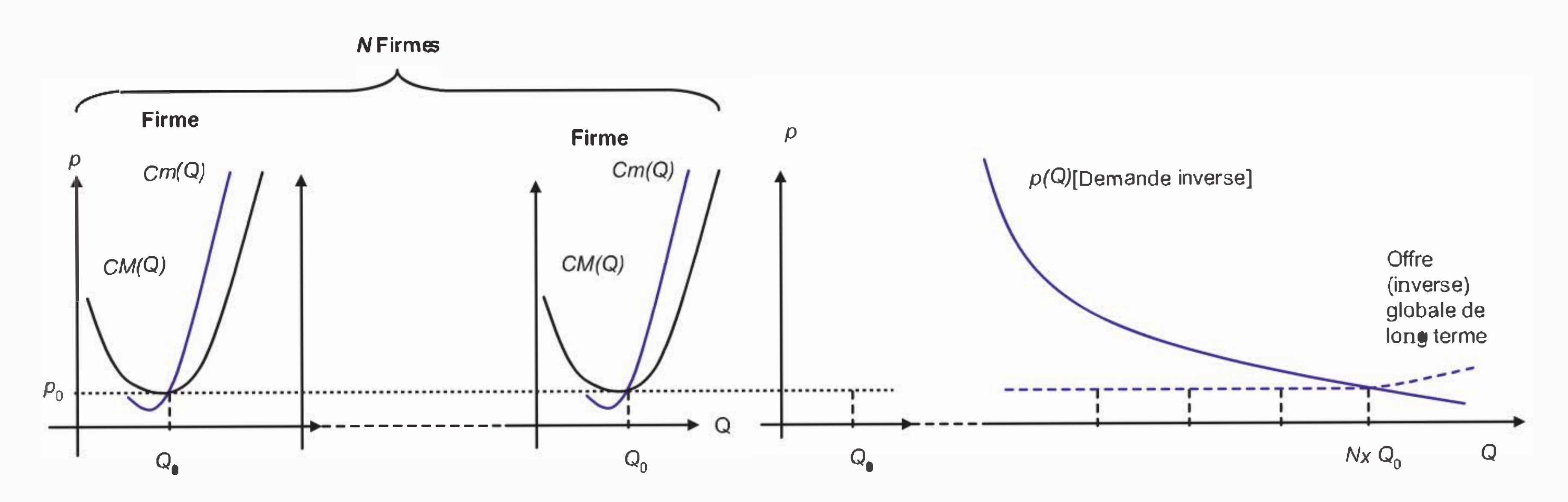


Figure 3 Offre globale de long terme et demande

Dans le cadre du modèle de concurrence pure et parfaite, **le long terme** désigne le délai à l'issue duquel :

1) Toutes les firmes qui produisent le bien ou service ont adopté la technologie de production la plus efficace disponible,

2) Le nombre de firmes présentes s'est ajusté de manière à ce que le profit microéconomique de chacune d'entre elles soit nul.

Il nous est aussi maintenant possible de donner la définition de l'équilibre (partiel) du marché en question.

On dit d'un marché qu'il est à **l'équilibre** lorsque les quantités offertes et les quantités demandées sur ce marché sont égales. Le prix qui conduit à ce que les quantités offertes et demandées soient ainsi égales est appelé prix d'équilibre.

Nous avons maintenant une vue d'ensemble de l'édifice théorique qu'est le modèle de concurrence pure et parfaite. Nous comprenons, en particulier, pourquoi chaque entrant adopte nécessairement la technologie la plus performante : si, dans le délai qui sépare la situation initiale (où seul un petit nombre de firmes technologiquement hétérogènes sont présentes) et le « long terme » un entrant choisissait une technologie autre que la plus efficace, il aurait, peut-être, temporairement, la possibilité de produire et vendre sans subir de perte, mais l'abaissement progressif du prix vers sa valeur plancher (le minimum du coût moyen supporté par toute firme adoptant la technologie la plus efficace) l'éjecterait nécessairement du marché avant que le prix ne se stabilise au prix plancher. Le processus d'ajustement du nombre de firmes présentes sur le marché est un processus qui joue à la hausse comme à la baisse. Si, à court terme, il y a plus de firmes que le nombre conduisant à l'équilibre de long terme entre offre et demande (même si elles sont toutes technologiquement parfaitement performantes), certaines disparaîtront, de manière *a priori* aléatoire.

Proposition Équilibre d'un marché en concurrence pure et parfaite

À l'équilibre de concurrence pure et parfaite, N firmes recourant à une même technologie (la plus efficace) vont chacune produire une quantité d'output Q^* telle que cette production se fasse au minimum de leur coût unitaire (commun) de production. Le prix d'équilibre p^* sera égal au minimum de ce coût unitaire. A ce prix p^*, la demande globale exprimée par les consommateurs sera précisément égale à $N \times Q^*$. Ainsi, les quantités globalement offertes et demandées sont identiques.

La situation d'équilibre du marché de concurrence pure et parfaite est représentée sur la figure 4.

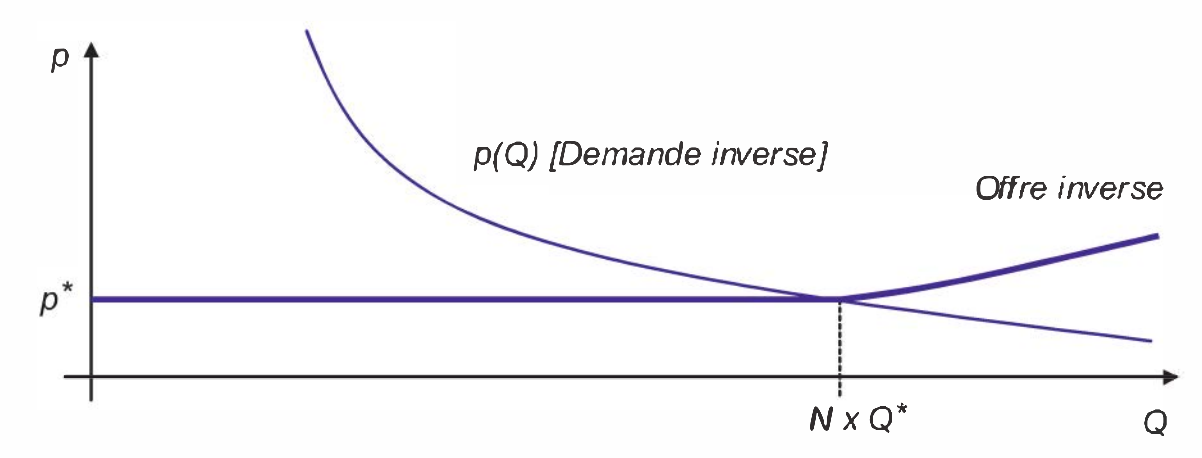


Figure 4 Équilibre du marché en concurrence pure et parfaite

Le modèle de concurrence pure et parfaite est une construction théorique qui semble assez lointaine de la réalité car de nombreuses hypothèses sont trop éloignées de ce que l'on observe effectivement dans les véritables secteurs d'activité : homogénéité du bien, libre entrée (et sortie), atomicité des firmes, etc. Mais ce modèle a une grande vertu : il est un **étalon de l'efficacité économique**. En parvenant, à long terme, à une situation dans laquelle la production se fait spontanément au minimum du coût unitaire (et par un nombre optimal de firmes), l'économie aboutit à **l'efficacité productive** : en aucun cas, au regard de la technologie disponible, il n'aurait été possible de produire une même quantité d'output pour un coût global inférieur. Ce résultat est le complément du résultat énoncé dans le

chapitre 5, où nous montrions qu'à l'équilibre d'une économie d'échanges purs, **l'efficacité allocative** était atteinte. Il permet de comprendre de manière plus aboutie le premier théorème de l'économie du bien-être (énoncé à la fin du chapitre 5) : dans une économie constituée de plusieurs marchés, tous fonctionnant selon les modalités de la concurrence pure et parfaite, si les consommateurs ont des préférences strictement convexes (et en l'absence d'externalités), le libre fonctionnement des marchés et la libre opportunité d'échanger conduiront à une allocation des ressources optimale au sens de Pareto. Il est possible (nous ne le ferons pas dans ce manuel) de construire un modèle d'équilibre général avec production : il s'agirait de combiner la construction présentée dans le chapitre 5 (la manière dont les agents procèdent librement à l'échange de façon à accroître conjointement leurs satisfactions) et la construction présentée ci-dessus (la modélisation de la production des biens et services vendus sur des marchés parfaitement concurrentiels). Dans un tel modèle d'équilibre général avec production, il serait alors possible de mettre en évidence l'obtention d'un équilibre général constituant une allocation pleinement efficace des ressources.

Dans la prochaine section, nous allons délaisser l'ambitieuse construction de l'équilibre général pour nous focaliser sur les caractéristiques de l'équilibre sur un marché quelconque, considéré isolément. C'est ce que nous allons étudier sous la dénomination « équilibre partiel » (par opposition à « équilibre général »).

2 L'équilibre partiel d'un marché quelconque

Dans cette section, nous examinons le marché d'un bien ou service quelconque en « fermant les yeux » sur la formation des fonctions d'offre et de demande. Nous supposons simplement que de telles relations entre prix et quantités offertes et demandées existent et nous nous contentons de caractériser le couple prix-quantité qui se forme à l'équilibre du marché. Néanmoins, nous faisons l'hypothèse implicite que les offreurs et les demandeurs sont suffisamment « petits » et nombreux pour ne pas

être en mesure, individuellement, de peser sur la formation du prix. Sans nous inscrire sur un marché de concurrence pure et parfaite, nous nous plaçons donc dans l'optique d'un marché « concurrentiel ».

Pour faciliter le raisonnement, nous allons désormais faire figurer le prix en abscisse et les quantités en ordonnée (contrairement à ce qui apparaissait dans la section précédente) ; ainsi ce sont désormais les offres et demandes « directes » – et non inverses – qui figureront dans les graphiques. Sur la figure 5, nous faisons apparaître une fonction de demande décroissante et une fonction d'offre croissante. Seule la configuration de concurrence pure et parfaite conduit à envisager que la fonction d'offre soit d'abord constituée d'une droite verticale, prolongée par une courbe fortement croissante à partir d'une quantité seuil correspondant au volume d'output réalisant l'équilibre du marché.

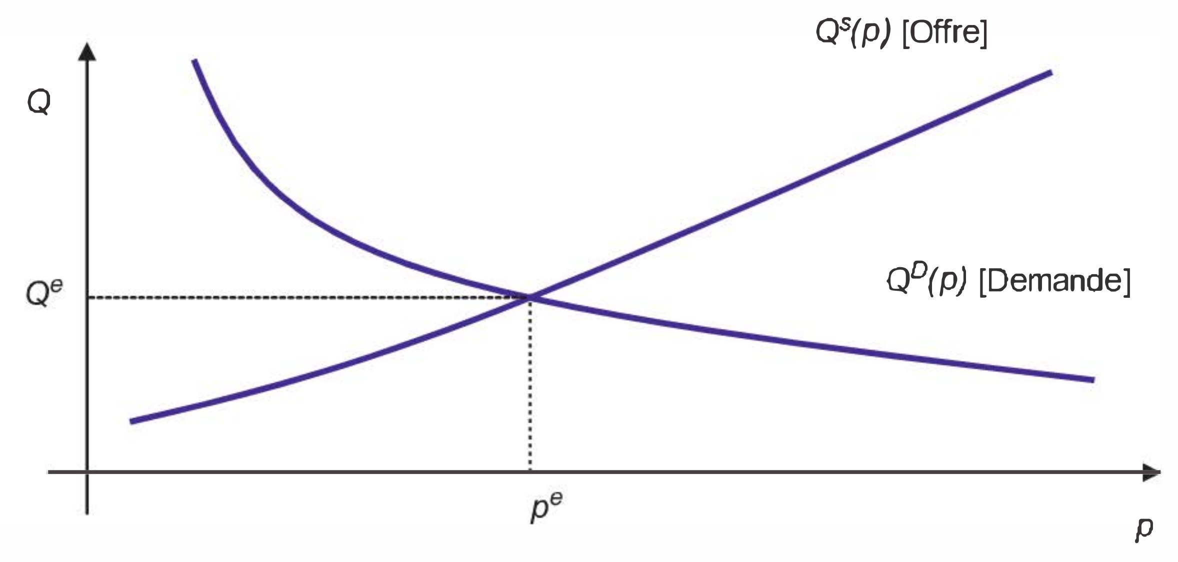


Figure 5 Fonctions d'offre et de demande et équilibre du marché

Sur cette figure 5, le couple prix-quantité réalisant l'équilibre sur le marché est le couple (p^e; Q^e). Que signifie cette proposition ? Pour bien en saisir le sens, il faut revenir à la définition première de ce qu'est un équilibre.

Un **système** est à l'**équilibre** s'il a atteint une situation de laquelle il ne s'écarte pas.

Cette définition rappelle beaucoup la définition de l'équilibre général d'une économie d'échange donnée dans le chapitre 5. Dans le cas où le système étudié est le marché d'un bien ou service quelconque, l'équilibre du marché, nous l'avons dit plus haut, est une situation où les quantités offertes et les quantités demandées sur ce marché sont égales. Être à l'équilibre signifie que chaque agent (producteur ou consommateur) choisit l'action qu'il préfère et que son comportement est compatible avec celui de tous les autres. *A contrario*, à tout prix différent du prix d'équilibre, les décisions que chacun des agents souhaiteraient prendre ne sont pas toutes réalisables. C'est donc le prix qui joue le rôle de signal, de vecteur de l'information.

Ce qui n'est pas explicite ici est la manière dont la coordination entre les producteurs et les consommateurs se fait, ou, en d'autres termes, la manière dont le prix parvient au niveau qui égalise l'offre et la demande. Dans le chapitre 5, nous avons évoqué l'idée d'un commissaire priseur qui collecte, pour chaque niveau de prix, l'information sur les intentions d'offres et de demandes des acteurs du marché et parvient, en tâtonnant, à établir le prix qui égalise parfaitement les offres et les demandes. Si ceci existe vraiment sur quelques marchés locaux ou mondiaux pour des denrées alimentaires, des fleurs, des métaux ou des pierres précieuses, il ne peut en être ainsi pour l'ensemble des biens et services. Cependant, par un étrange méandre de l'Histoire, il paraît désormais concevable qu'un site internet consulté et utilisé en temps continu sur la surface entière du globe puisse, en permanence, enregistrer les intentions d'offres et de demandes de chaque agent économique et fixer un prix d'équilibre pour chaque bien ou service. En vérité, ce type de dispositifs existe déjà. Ainsi, en effet, certains actifs financiers sont côtés en temps continu, mondialement. Dans le cas de biens et services plus locaux, le mécanisme d'ajustement du prix est plus « terre à terre », fondé sur le bon sens et l'expérience des acteurs. Ainsi, par exemple, un maraîcher qui vend des légumes sur une place de marché, ajustera, en cours de matinée, le prix de ses produits : s'il voit que ses concurrents ont de faibles volumes à proposer et/ou s'il a déjà écoulé les trois quarts de sa production à mi-« séance », il a tendance à augmenter ses prix ; à l'inverse, si la ressource est très abondante sur tous les étals et/ou si son stock peine à « partir », il a intérêt

à baisser drastiquement ses prix pour ne pas repartir avec ses invendus périssables.

L'une des questions qui se pose quant à la détermination du prix d'équilibre est l'impact d'un accroissement soit de l'offre, soit de la demande, à partir d'une situation initiale d'équilibre (que nous désignerons par $(p_0 ; Q_0)$).

2.1 Impact d'un accroissement de l'offre

Supposons que, suite à une innovation technologique ou organisationnelle, une ou plusieurs firmes soient désormais en mesure d'offrir, pour chaque niveau de prix, une quantité plus grande du bien. Il en résulte un **déplacement** vers le haut **de la courbe** d'offre. Le prix d'équilibre (initialement p_0), s'abaisse progressivement jusqu'au niveau p_1 qui réalise le nouvel équilibre entre les quantités offertes et demandées (figure 6). Concrètement, les firmes qui ont choisi de mettre en œuvre l'innovation produisent et vendent plus d'unités qu'auparavant (mais il n'est pas sûr que leur profit croisse car le prix de vente s'abaisse) ; les firmes restées passives – en matière technologique ou organisationnelle – voient vraisemblablement leur profit baisser (mais tout dépend de l'impact respectif de la hausse ou de la baisse de leur coût unitaire, relativement au volume produit et à la baisse du prix). Quant aux consommateurs, ils bénéficient collectivement de cette baisse du prix : des consommateurs dont la disponibilité maximale à payer était auparavant trop faible – ceux dont la disponibilité maximale à payer est comprise entre p_0 et p_1 – peuvent désormais consommer ce bien et, à l'exception du consommateur marginal (celui dont la disponibilité maximale à payer est exactement égale au nouveau prix p_1), ils réalisent un gain en bien-être. Cet élargissement du socle des agents consommateurs du bien correspond au **déplacement** vers la gauche **sur la courbe** de demande qui conduit au nouvel équilibre (de $(p_0 ; Q_0)$ à $(p_1 ; Q_1)$) (remarquons que ceci correspond à un déplacement vers la droite sur la courbe de demande inverse, ce qui, en matière de visualisation de l'accroissement du surplus des consommateurs est, sans doute, plus intuitif).

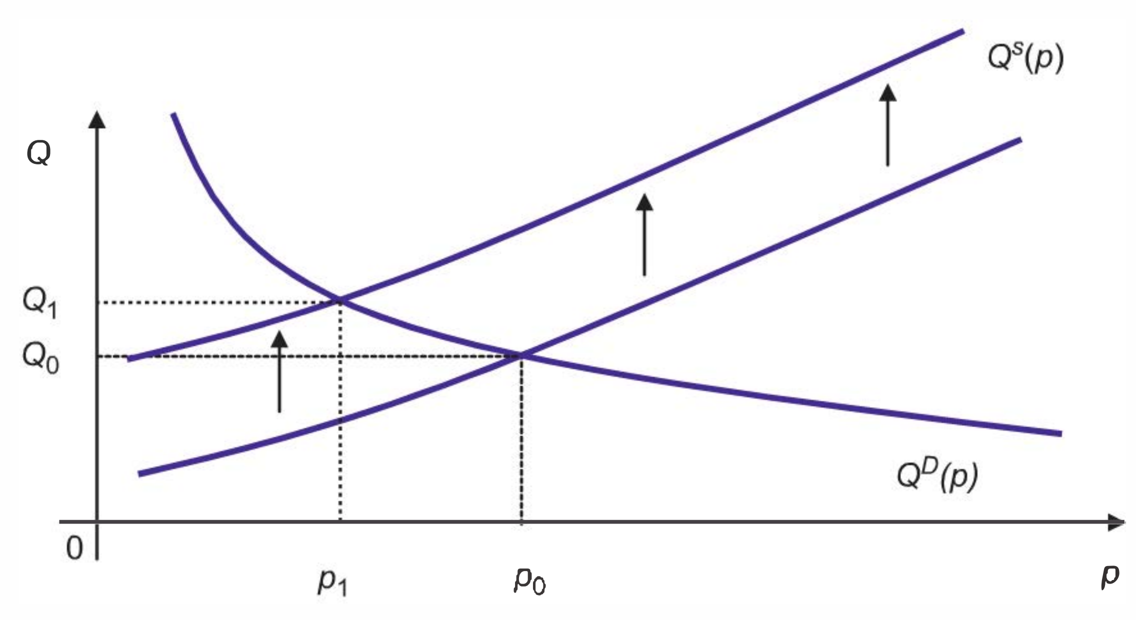


Figure 6 Impact d'un accroissement de l'offre

2.2 Impact d'un accroissement de la demande

Supposons que, suite à une progression démographique, et/ou suite à la suppression de barrières douanières, et/ou suite à l'ouverture d'une nouvelle desserte (maritime, routière ou fluviale) et/ou suite à l'apparition d'un ou plusieurs nouveaux distributeurs, etc., la quantité demandée du bien soit, pour chaque niveau de prix, plus conséquente. Il en résulte un **déplacement** vers le haut **de la courbe** de demande. Le prix d'équilibre (initialement p_0), augmente progressivement jusqu'au niveau p_2 qui réalise le nouvel équilibre entre les quantités offertes et demandées (figure 7). Pratiquement, la compétition accrue entre les consommateurs pour l'obtention des unités du bien conduit à ce que seuls ceux dont la disponibilité maximale à payer est supérieure ou égale à p_2 continuent à participer à la consommation du bien. On enregistre donc une perte de bien-être pour ceux des consommateurs, initialement présents, qui continuent de participer à la consommation du bien (car chaque unité consommée est payée plus chère qu'auparavant) et pour tous ceux qui sont désormais exclus de la consommation du bien. À l'inverse, certains consommateurs connaîssent un gain en bien-être : ceux des consommateurs nouvellement desservis (si nous supposons, pour fixer les idées, que l'accroissement de la demande se fait par l'extension du champ géographique sur lequel le bien ou service est distribué) qui ont une disponibilité maximale à payer supérieure ou égale à

8

p_2 réalisent un gain en bien-être. Quant aux firmes, elles bénéficient, sans équivoque, collectivement de cette hausse conjointe du prix et des quantités produites. L'accroissement des quantités produites se traduit par un **déplacement** vers la droite **sur la courbe** d'offre qui conduit au nouvel équilibre (de $(p_0; Q_0)$ à $(p_2; Q_2)$).

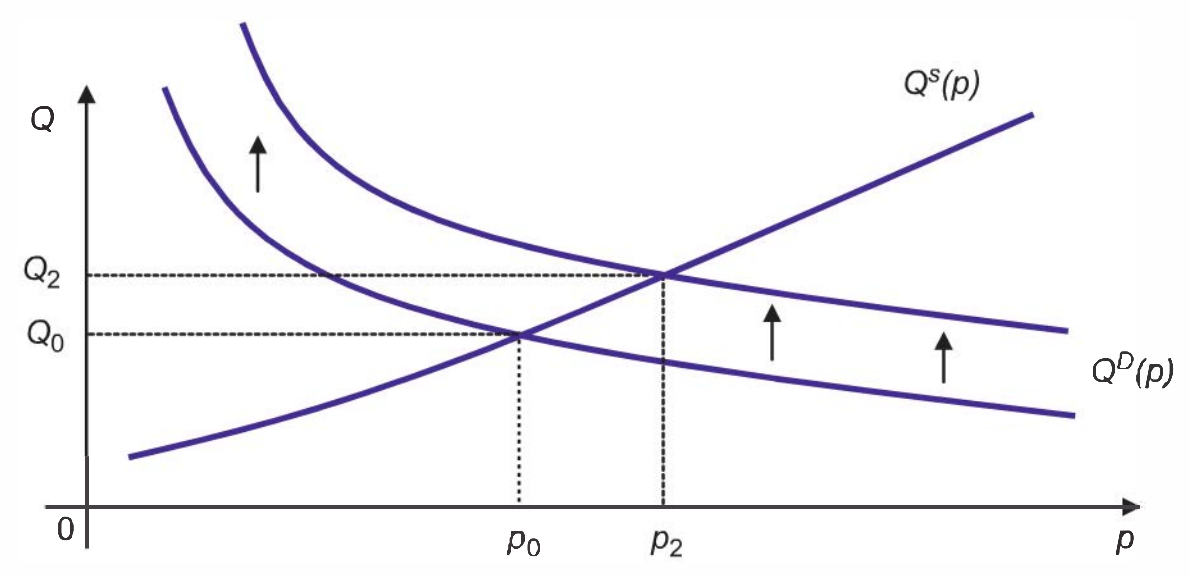


Figure 7 Impact d'un accroissement de la demande

Les raisonnements ci-dessus pourraient, bien entendu, être développés en sens inverse à l'occasion de l'étude de l'impact d'une contraction de l'offre ou de la demande. Il est essentiel de ne pas confondre les mouvements des courbes et les mouvements le long des courbes. Les déplacements des courbes elles-mêmes sont causés par des modifications des conditions de l'offre ou de la demande : changement technologique, élargissement ou rétrécissement des débouchés. Les déplacements le long des courbes sont, eux, la réponse des firmes ou des consommateurs à une variation du prix auquel les échanges peuvent se produire.

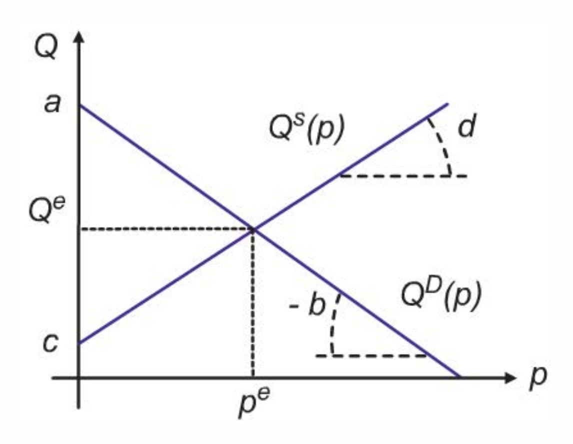
Dans la dernière section, nous allons tenter d'approfondir, à fonctions d'offre et de demande données, le scénario de cheminement vers le prix d'équilibre. Nous évoquerons alors la notion de stabilité de l'équilibre.

Pour aller plus loin

Détermination du prix et de la quantité d'équilibre dans le cas de fonctions d'offre et de demande linéaires

Supposons que la demande globale pour le bien soit : $Q^D(p) = a - bp$ et que l'offre globale pour le bien soit : $Q^s(p) = c + dp$, où a, b, c et d sont des paramètres strictement positifs (et tels que $a > c$). Le prix d'équilibre du marché est le prix pour lequel la quantité totale offerte et la quantité totale demandée sont égales : $Q^d(p) = Q^s(p) \Leftrightarrow p^e = \frac{a-c}{b+d}$

À ce prix, la quantité offerte comme la quantité demandée sont égales à $Q^e = \frac{ad+bc}{b+d}$



3 Cheminement vers l'équilibre : le modèle du cobweb

Nous l'avons dit, le principe d'un cheminement immédiat vers l'équilibre est peu réaliste. Pour se rapprocher de la réalité, on peut tenter de modéliser un cheminement vers l'équilibre en ne niant pas que ce processus prenne du temps. Un secteur d'activité se prête particulièrement bien à la modélisation du processus d'ajustement temporel vers l'équilibre : c'est le

secteur de la production agricole. En effet, lorsque les producteurs décident de la surface à cultiver d'une céréale ou d'un légume particulier, ils le font sur la base de l'observation des cours aujourd'hui. Mais demain, quand les céréales auront poussé, les cours seront peut-être différents et la demande le sera aussi. On devine que si les cours ont monté, la demande sera plus faible que ce qu'ils avaient anticipé à la période précédente et, qu'en conséquence, leur production effective va excéder la demande. À l'inverse, si les cours ont baissé, la demande sera plus forte que ce qu'ils pensaient et ils ne disposeront pas d'une production suffisante pour satisfaire leurs clients.

Formalisons cette idée : supposons que les cultivateurs, sur la base de l'observation du prix à la date t, décident de la surface de céréale à mettre en culture. On considère que cette décision permet de déterminer, avec précision, la quantité totale de céréales qui sera mise en vente à la date $t + 1$. On peut ainsi écrire que l'offre des firmes (les cultivateurs) est une fonction $Q^S_{t+1} = f(p_t)$: la quantité produite à la date $t + 1$ est une fonction (croissante) du prix à la date t. En revanche, la quantité demandée dépend du prix observé lors de la période présente. Ainsi, $Q^D_t = g(p_t)$ (où g désigne une fonction décroissante). En d'autres termes, s'il n'y pas de décalage temporel du côté de la demande, il y en a un du côté de l'offre.

Partons d'une situation initiale d'équilibre, que nous désignons par $(p_0 ; Q_0)$, et supposons qu'un événement extérieur vienne perturber le système : ici, nous raisonnons sur les conséquences d'un choc négatif sur l'offre. Par exemple, des intempéries dévastatrices (orages de grêles) déciment les récoltes. En conséquence, la production effective Q_1 sera beaucoup plus faible que Q_0. Le prix ne peut pas demeurer égal à p_0 car la quantité demandée serait alors trop importante au regard de la quantité offerte ; les cultivateurs vont donc accroître leur prix de manière à ce que la demande exprimée redescende au niveau de l'offre disponible. Ainsi, le prix exigé va croître du niveau p_0 au niveau p_1. C'est sur la base de ce nouveau prix p_1 qu'ils vont décider de la surface de céréales à mettre en culture. En l'absence de nouvelles intempéries, cette décision va conduire à une quantité offerte (en période suivante) Q_2 qui sera, cette fois, excessive au regard de la demande qui s'exprimerait si le prix demeurait égal à p_1. Pour ne pas rester « à la tête » de quintaux invendus, les agriculteurs vont donc consentir à abaisser leur prix, jusqu'au niveau p_2, prix sur la base duquel

sera décidée la quantité Q_3 produite la saison suivante, et ainsi de suite… Le processus en question est représenté sur la figure 8. On observe un mouvement de convergence vers le prix (et la quantité) d'équilibre. Dans cette configuration, la position relative des courbes d'offre et de demande est telle que, quelle que soit la perturbation initiale, les forces du système le ramèneront à sa position d'équilibre. Il s'agit d'un équilibre dit stable.

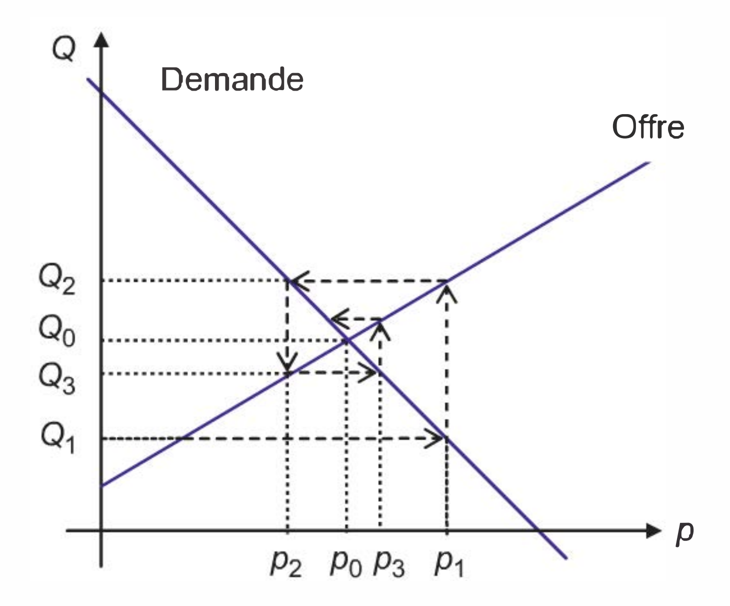


Figure 8 Cheminement vers l'équilibre du marché

Le schéma sur la figure 8 ressemble à une toile d'araignée. C'est pourquoi on présente cette modélisation comme le modèle du « cobweb ». Comme indiqué ci-dessus, dans le cas présenté sur la figure 8, le processus est convergent ou implosif : l'équilibre est stable.

Un **équilibre** est dit **stable** si, suite à une perturbation du système, les forces du système le ramènent à sa position initiale.

Il se peut, à l'inverse, que les caractéristiques de l'offre et de la demande fassent de tout équilibre une situation instable : toute perturbation, aussi minime soit elle, affectant une situation initialement à l'équilibre va entraîner un processus qui éloigne le système de sa position d'équilibre. On parlera d'un cobweb divergent ou explosif. Une telle situation est représentée sur la figure 9.

8

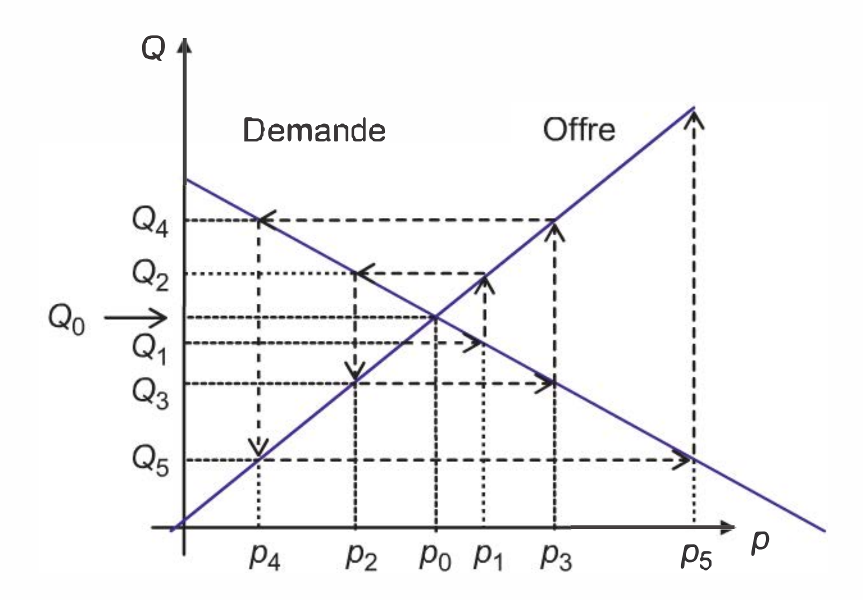


Figure 9 Cobweb explosif

On peut deviner graphiquement que la stabilité du système est garantie si la valeur absolue de la pente de la courbe de demande est plus forte que la pente de la courbe d'offre. Plus exactement :

- Si l'élasticité-prix de la demande (en valeur absolue) est plus forte que celle de l'offre, nous serons confrontés à un cobweb implosif, c'est-à-dire un processus dans lequel les écarts de prix sont de plus en plus faibles au fur et à mesure des périodes, et qui, à l'issue d'un certain délai, ramène le marché à l'équilibre.
- Si l'élasticité-prix de la demande (en valeur absolue) est plus faible que celle de l'offre, nous serons face à un cobweb explosif, c'est-à-dire un processus dans lequel les écarts de prix s'accroissent au fur et à mesure des périodes, et qui éloigne le marché de sa position d'équilibre.

Nous avons décrit, dans ce chapitre, les caractéristiques de l'équilibre d'un marché comportant implicitement un grand nombre de firmes et de consommateurs. Dans les deux chapitres suivants, nous conserverons l'hypothèse d'une multitude de consommateurs, mais nous envisagerons désormais des marchés sur lesquels ne sont présents qu'un petit nombre de firmes. Le premier de ces deux chapitres traite du cas extrême de cette réduction du nombre de producteurs : c'est le cas du monopole.

9 Le monopole

Mots-clés

Pouvoir de marché, monopole naturel, tarifications de 1er et de 2e rang, discrimination tarifaire

Dans le chapitre précédent, nous avons étudié la décision de firmes si nombreuses et de si petites tailles qu'elles n'avaient aucune influence sur le niveau auquel se fixait le prix auquel il leur était possible de vendre leur output. Il s'agissait du cadre hypothétique de la concurrence pure et parfaite où les firmes sont *price takers*. Aux antipodes, nous allons maintenant étudier le cadre dans lequel l'offreur est en position de force absolue car seul et unique offreur de l'output. De manière assez évidente, cette position d'unique offreur fera de lui un *price maker*, un faiseur de prix.

En réalité, établir qu'une firme est (ou n'est pas) en position de monopole sur un marché n'est pas toujours très aisé. Si, dans une ville ou une communauté de communes, une firme est le seul distributeur d'eau potable par un unique réseau, il est possible d'affirmer, sans réserve, que cette firme est en situation de monopole. En revanche, lorsque l'on considère une firme qui produit et vend un vin d'exception, comme le château Haut Brion, doit-on considérer qu'elle est en position de monopole (il n'y a qu'un seul et unique château Haut Brion) ou est-il plus pertinent de considérer que Haut Brion est sur un marché en concurrence avec d'autres vins de Pessac Léognan ou même d'autres grands vins de Bordeaux ? La frontière entre une situation de monopole et une position où la firme est en concurrence avec d'autres firmes est parfois ténue. Dans ce chapitre, nous nous contenterons de supposer que la firme dont on étudie la décision est effectivement en situation de monopole.



1 Monopole : définition et origine

Un **monopole** est un secteur d'activité où une firme unique, appelée monopole, est le seul offreur du bien ou service produit et mis en vente.

Quelles peuvent être les origines du monopole ? On distingue, souvent, deux origines possibles à l'existence d'un monopole sur un marché : une origine institutionnelle et une origine « économique ». Ces deux origines ne sont pas exclusives. Souvent même, un monopole institutionnel est instauré en raison de l'origine « économique » de ce monopole. Indiquons ce que recouvrent ces deux notions.

- **Origine institutionnelle** : la législation ou la réglementation ont institué l'existence d'un monopole. Soit l'activité n'existait pas et la puissance publique a créé de toutes pièces une structure nouvelle, soit elle a regroupé toutes les structures existantes, soit elle a décrété que l'unique firme présente devait demeurer seule productrice de l'output. Dans certains cas, la création d'un monopole institutionnel nécessite une (ou des) nationalisations.
- **Origine « économique »** : situation dans laquelle il est économiquement justifié et rationnel qu'une firme unique produise la totalité des unités proposées à la vente. On parle alors d'un « **monopole naturel** ». Très souvent, les monopoles naturels sont contrôlés par la puissance publique (ils deviennent également, alors, des monopoles institutionnels).

L'activité de distribution d'eau potable dans un bassin économique (une ville) est un monopole naturel. Sans entrer dans les détails précis (cf. section 4), chacun peut concevoir qu'il serait absurde de constituer deux réseaux parallèles de distribution d'eau potable dans la ville. Un seul réseau est déjà très coûteux à édifier et entretenir, il serait économiquement inepte d'en bâtir deux (ou plus !). Il en va de même pour différentes activités ayant en commun de nécessiter la présence d'un réseau (de fils, de canalisations, de voies ferrées, de guichets, etc.). De tels monopoles naturels sont, de manière quasi systématique, contrôlés ou encadrés par la puissance publique (l'État ou les collectivités territoriales) : soit en faisant de la firme

une structure pleinement publique (par exemple une régie municipale de distribution d'eau), soit en délégant à une structure privée, par appel d'offres, la charge d'assurer l'activité en question, en respectant des règles très précises (en particulier de tarification).

2 Le comportement optimal du monopole non contraint

Nous allons maintenant examiner la décision d'un monopole quelconque (non nécessairement « naturel »), lorsqu'il n'est pas contraint, c'est-à-dire lorsqu'il n'est pas tenu de respecter des règles de tarification ou des quotas de production. Précisons en outre que le monopole ici étudié est non discriminant, c'est-à-dire qu'il vend chaque unité du bien ou service à un même prix unitaire. Nous évoquerons, dans la section 5, la possibilité de pratiquer une « discrimination » par les prix, c'est-à-dire commercialiser les différentes unités du bien produit à des prix différenciés.

Comme nous l'avons dit en introduction, le monopole possède le privilège de parfaitement connaître la relation entre prix et quantité demandée du bien ou service produit, sous forme, par exemple, de la fonction de demande inverse $p(Q)$; seul face à l'ensemble des consommateurs, il est capable de mesurer la sensibilité de la demande de ces consommateurs aux variations du prix du bien ou service qu'il commercialise. Il n'est donc plus « price-taker » mais bien « price-maker ». Cette caractéristique va avoir une conséquence décisive : la recette marginale tirée de la vente de l'output ne sera plus égale à son prix : elle sera inférieure. Car, au regard d'une situation donnée, si le monopole désire vendre des unités supplémentaires de bien, il doit proposer ces unités supplémentaires à un prix plus faible : pour trouver de nouveaux acheteurs, il doit se tourner vers des acquéreurs dont la disponibilité maximale à payer est inférieure au prix initial. Comme, en outre, la discrimination tarifaire est (par hypothèse) impossible, cela signifie que chaque unité du bien ou service (y compris celles vendues aux consommateurs initiaux) devra être vendue à ce nouveau prix unitaire plus faible.

Pour aller plus loin

À quoi est égale la recette marginale du monopole ?

Puisque la firme est faiseuse de prix ou « price maker », le prix unitaire p est désormais, non plus une constante, mais une fonction de la quantité vendue (la demande inverse, précisément) dans le calcul de la dérivée première de la fonction de recette totale. Ainsi, la recette totale de la firme en situation de monopole est :

$$RT(Q) = p(Q) \times Q$$

Si l'on calcule la recette marginale, il vient :

$$Rm(Q) = \frac{\partial RT(Q)}{\partial Q} = \frac{\partial(p(Q)\times Q)}{\partial Q} = \frac{\partial p(Q)}{\partial Q} \times Q + p(Q) \times 1$$

$$Rm(Q) = \frac{\partial p(Q)}{\partial Q} \times Q + p(Q)$$

Or le terme $\partial p(Q)/\partial Q$ est négatif (la demande est une fonction décroissante du prix). Donc $Rm(Q) < p(Q)$, la recette marginale est inférieure au prix.

La condition qui caractérise la décision optimale de toute firme désireuse de maximiser son profit microéconomique est l'égalité entre la recette marginale et le coût marginal (cf. chapitre 7). Pour représenter graphiquement la décision optimale d'un monopole (non contraint, non discriminant), il est indispensable de préciser la forme de la fonction de recette marginale, c'est-à-dire de la relation qui mesure, pour toute quantité vendue d'output, le surcroît de recette obtenu de la vente de la dernière unité. Le monopole connaît parfaitement la fonction de demande inverse et est donc capable de calculer la recette totale qui sera tirée de la vente de toute quantité d'output. La recette marginale est la dérivée première de la fonction de recette totale, qui prend la forme d'une fonction décroissante, située sous la fonction (elle-même décroissante) de demande inverse. La recette marginale est décroissante car pour vendre des unités d'output supplémentaires, elle devra toujours se tourner vers des consommateurs dont la disponibilité maximale à payer est toujours plus faible (la recette marginale est donc inférieure au prix et décroissante). On a pour coutume de représenter la recette marginale et la

demande inverse en désignant la demande inverse sous l'appellation « recette moyenne ». Ceci a la vertu de permettre à la paire *Recette marginale - Recette Moyenne* de « répondre » à la paire *Coût marginal - Coût Moyen*. Précisons ce qu'est la recette moyenne $RM(Q)$.

La **recette moyenne** tirée de la vente d'un bien ou service, notée $RM(Q)$ est le quotient de la recette totale $RT(Q)$ sur la quantité. Elle mesure quelle recette est tirée, en moyenne, de la vente d'une unité de ce bien ou service.

$$RM(Q) = \frac{RT(Q)}{Q}$$

Par construction, pour un monopole, la recette moyenne est confondue avec la demande inverse. En effet, pour un monopole :

$$RT(Q) = p(Q) \times Q \text{ donc } RM(Q) = \frac{RT(Q)}{Q} = \frac{p(Q)\times Q}{Q} = p(Q)$$

Représentons, sur la figure 1, les recettes moyenne $RM(Q)$ et marginale $Rm(Q)$.

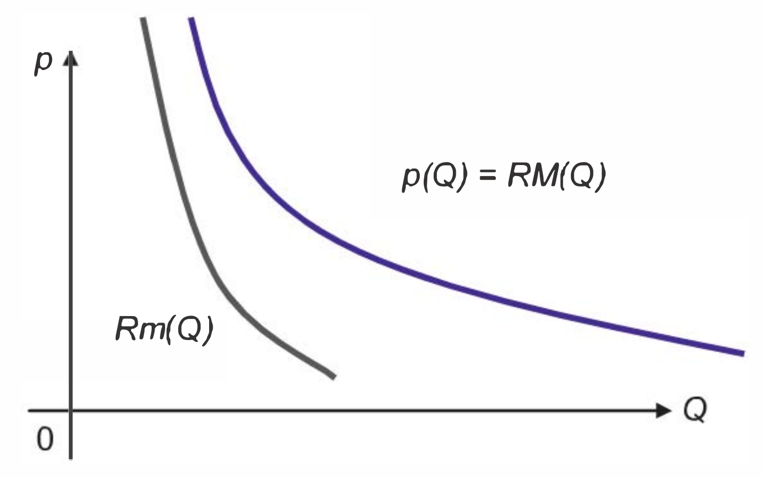


Figure 1 Recette moyenne et recette marginale

Pour figurer la décision optimale du monopole, il faut maintenant faire apparaître la fonction de coût marginal, puisque la quantité Q^M qui égalise la recette marginale et le coût marginal est celle qui permet au monopole de maximiser son profit. Outre la détermination de la quantité optimale, le

monopole s'efforce d'établir le prix optimal auquel il a intérêt à mettre en vente ces unités d'output, car c'est bien d'un couple prix-quantité dont il décide. Sous les hypothèses détaillées ici, établir le prix optimal est une tâche mécanique : il suffit de « lire » ce prix en observant la courbe de demande inverse (ou, ce qui est équivalent ici, la courbe de recette moyenne). En effet, dès lors que la quantité optimale à produire a été déterminée, seul le prix mécaniquement attaché à cette quantité par la relation de demande inverse est celui qui assure à l'entrepreneur l'obtention d'un profit maximum. Cette quantité Q^M et ce prix p^M optimaux sont identifiés sur la figure 2.

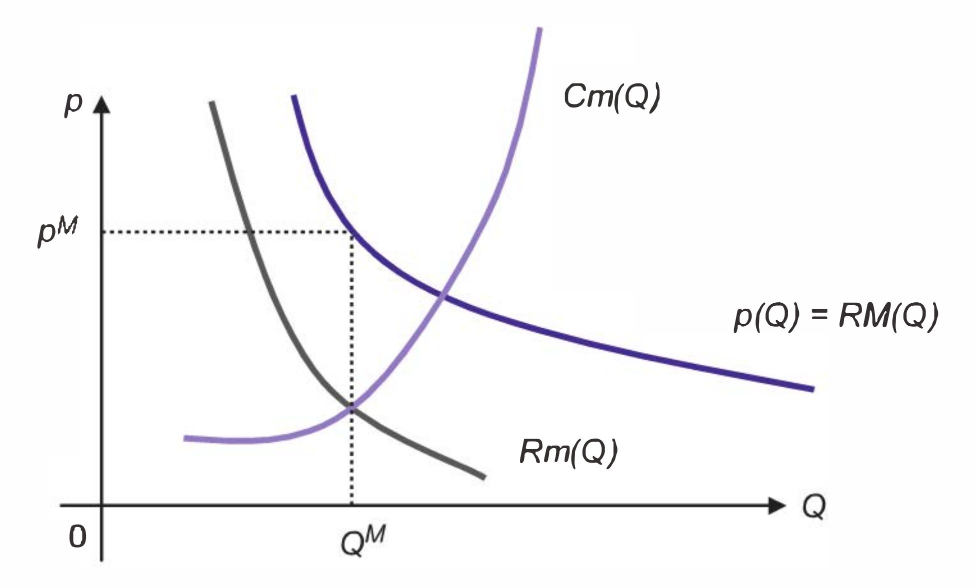


Figure 2 Décision optimale du monopole non contraint

Sur la figure 2 est représentée une courbe de coût marginal croissante. Ceci est une option qui ne correspond pas nécessairement à ce que l'on trouve auprès de la grande majorité des firmes en situation de monopole. Il s'agirait ici, par exemple, d'une firme que la puissance publique a placée (et protège) institutionnellement en situation de monopole. Nous verrons, dans la section sur le monopole naturel, quelle forme ont les courbes de coût lorsque les circonstances technologiques « commandent » la présence d'un monopole. Dans la configuration représentée sur la figure 2, le prix que peut pratiquer le monopole non contraint est élevé. Ce prix est susceptible d'être et demeurer, à long terme, supérieur au coût moyen de production. Ainsi, le monopole est susceptible de réaliser durablement un profit microéconomique strictement positif, contrairement à ce qui se passe pour

une firme en situation de concurrence pure et parfaite. Il est possible de montrer que l'importance de ce profit est inversement proportionnelle à l'élasticité-prix de la demande, notée ε. En effet, si nous construisons un indicateur appelé « taux de marge », noté μ, qui mesure la capacité qu'a le monopole à vendre ses unités d'output à un prix élevé (au regard de son coût marginal de production), il est possible de mettre en évidence la propriété suivante :

$$\mu^M = \frac{p^M - Cm(Q^M)}{p^M} = -\frac{1}{\varepsilon} = \frac{1}{|\varepsilon|}$$

Le taux de marge est inversement proportionnel à (la valeur absolue de) l'élasticité-prix de la demande. Si la demande est très inélastique (gaz ou électricité pour le chauffage des logements), le prix auquel un monopole non encadré (non contraint) peut commercialiser son énergie est très élevé et le profit qu'il en tire aussi. À l'inverse, si la demande est très élastique, le prix que peut pratiquer le monopole est faible, à peine supérieur au coût marginal. Il en résulte l'obtention d'un profit quasiment nul. Ainsi, dans le cas où la demande est très élastique au prix, les conditions de commercialisation de l'output s'apparentent à celles d'une situation fortement concurrentielle, même si la production en est assurée par un monopole non encadré. Le taux de marge μ est interprété comme une mesure du **pouvoir de marché** du monopole.

Le **pouvoir de marché** d'une firme est sa capacité à réaliser un profit élevé sur ce marché. On le mesure par le taux de marge μ, qui est un indicateur de sa capacité à pratiquer durablement un prix supérieur à son coût marginal de production. De manière générale :

$$\mu = \frac{p - Cm(Q)}{p}$$

La définition ci-dessus présente un inconvénient : ce n'est pas parce que le prix pratiqué est supérieur au coût marginal qu'il est nécessairement supérieur au coût moyen. En conséquence, dans certaines situations, le taux de marge peut être positif alors même que le profit microéconomique est

négatif ! Néanmoins, l'écart entre le prix et le coût marginal demeure un « marqueur » pertinent de la capacité de la firme à extraire un profit conséquent, un profit « anormal » au regard des résultats caractéristiques rencontrés dans le cadre de la concurrence pure et parfaite.

3 Mise en évidence de l'inefficacité sociale du monopole non contraint

Dans le précédent chapitre, nous avons tenté d'approfondir le premier théorème de l'économie du bien-être qui dispose que l'équilibre général d'un système de marchés concurrentiels est socialement efficace. L'efficacité sociale conduit, en particulier, à ce que le bien-être des différents acteurs soit le plus élevé possible. Il est donc possible de juger du caractère socialement efficace de la situation sur un marché en calculant le niveau de bien-être globalement atteint et en examinant s'il a atteint sa valeur maximale. Le bien-être globalement atteint par les acteurs d'un marché (consommateurs et producteurs) est mesuré par le *welfare* (parfois appelé surplus global). Le *welfare* est la somme des surplus des consommateurs et des surplus des producteurs. La mesure du surplus des consommateurs est la somme des écarts entre les disponibilités maximales à payer des consommateurs et le prix effectivement payé (cf. chapitre 4). Quant au surplus des producteurs, il est simplement égal à la somme de leurs profits. En effet, il n'est nul besoin de chercher un indicateur de bien-être des firmes autre que le profit microéconomique : si, dans le cas des consommateurs, nous étions embarrassés pour construire un équivalent monétaire du bien-être que ceux-ci éprouvaient, dans le cas des producteurs, nous possédons, avec le profit, un indicateur qui permet une lecture directe de leur satisfaction en termes « monétaires ».

Le ***welfare*** est une mesure du bien-être éprouvé par une société. Il est concordant avec le critère d'efficacité économique au sens de Pareto. Le *welfare*, noté W, est égal à la somme des surplus des consommateurs et des producteurs, c'est-à-dire à la somme des surplus des consommateurs et des profits.

Dans le cas de l'étude d'un seul marché, le *welfare* relatif à ce marché est la somme du surplus des consommateurs engendré par la consommation du bien ou service produit sur ce marché et des profits (positifs ou négatifs) réalisés par l'ensemble des firmes qui produisent le bien ou service. Dans le cas du monopole, le *welfare* est donc égal à la somme du surplus des consommateurs et du profit du seul monopole : $W = SC + \Pi$. Il est possible de visualiser le niveau du *welfare* à partir d'un calcul d'aires. En effet, dans le cas où le coût fixe supporté par la firme est nul, le *welfare* est égal à l'aire située entre la courbe de demande inverse et la courbe de coût marginal. Ainsi, dans le cas où la décision du monopole est le couple $(p^M; Q^M)$, le *welfare* est la surface située, de 0 à Q^M, entre ces deux courbes (voir figure 3).

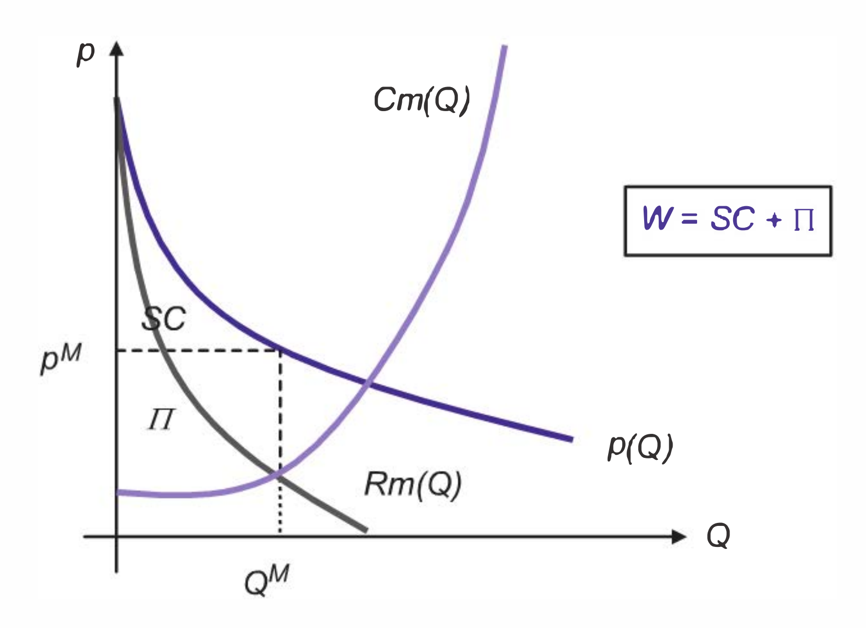


Figure 3 *Welfare* en présence d'un monopole (sous l'hypothèse d'un coût fixe nul)

Puisque la situation de concurrence pure et parfaite conduit au maximum d'efficacité sociale (maximum de *welfare*), visualiser la perte d'efficacité sociale (la perte de *welfare*) engendrée par l'existence d'un monopole peut se faire en comparant les *welfare* obtenus en configuration concurrentielle et en situation de monopole. Le couple prix-quantité qui correspondrait au fonctionnement parfaitement concurrentiel du marché se trouve à l'intersection entre la courbe de demande inverse et la courbe de coût marginal (cf. chapitre 8, le prix est précisément égal au coût marginal à l'équilibre concurrentiel). Le *welfare* qui serait obtenu sous l'hypothèse d'un fonctionnement concurrentiel du marché est représenté sur la figure 4.

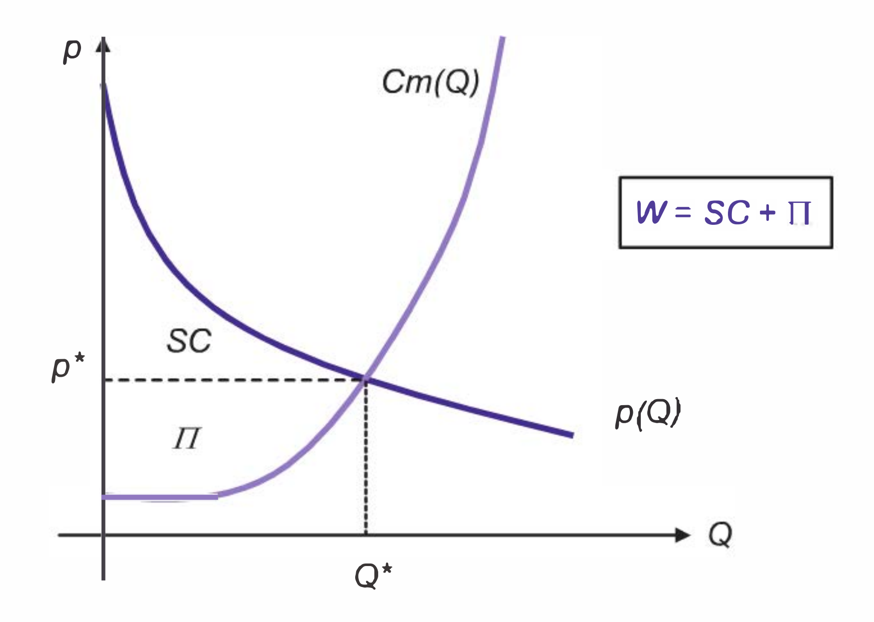


Figure 4 *Welfare* en contexte concurrentiel (sous l'hypothèse d'un coût fixe nul)

Pour aller plus loin

Représentation graphique du *welfare*

Comme nous l'avons indiqué dans le chapitre 4,

$$SC = \int_0^{Q^M} p(Q)dQ \; - p^M \times Q^M.$$

D'autre part, puisque la fonction de coût marginal est la dérivée de la fonction de coût total, la fonction de coût total est réciproquement la primitive de la fonction de coût marginal. Ainsi, $\Pi = p^M \times Q^M - C(Q^M)$ $= p^M \times Q^M - \int_0^{Q^M} Cm(Q)dQ$ et donc,

$$W = SC + \Pi = \int_0^{Q^M} p(Q)dQ \; - p^M \times Q^M + p^M \times Q^M - \int_0^{Q^M} Cm(Q)dQ$$

$$= \int_0^{Q^M} p(Q)dQ \; - \int_0^{Q^M} Cm(Q)dQ$$

Ainsi, $W = \int_0^{Q^M} [p(Q) - Cm(Q)]dQ$. Le *welfare* éprouvé, lorsque le couple prix-quantité est le couple $(p^M; Q^M)$, est égal à l'aire située, de 0 à Q^M, entre la courbe de demande inverse et la courbe de coût marginal.

En comparant les deux représentations, on visualise la perte de *welfare* ΔW concédée entre la situation concurrentielle et la situation de monopole. C'est ce qui apparaît sur la figure 5.

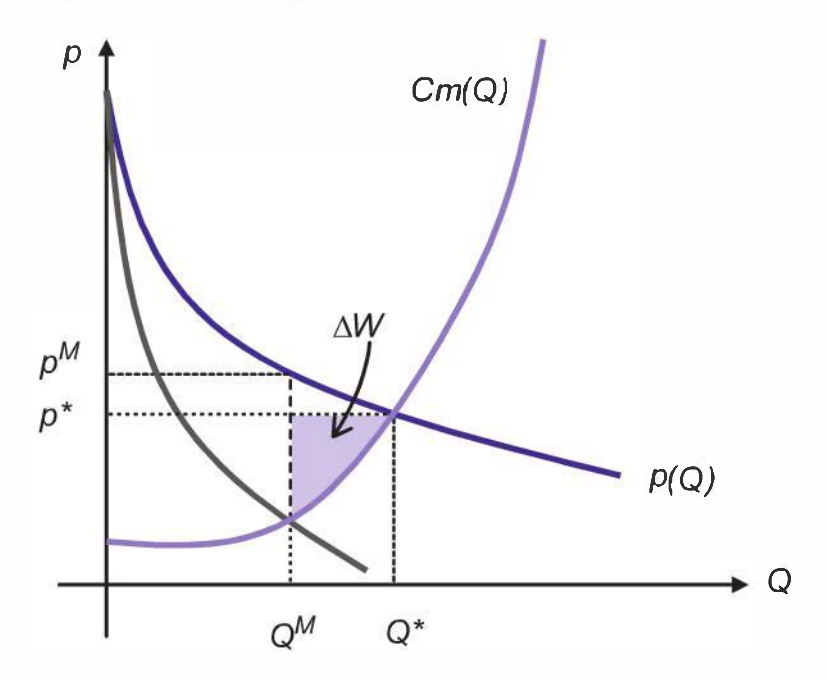


Figure 5 Perte de *welfare* liée à la présence d'un monopole (au regard d'un fonctionnement concurrentiel du marché)

La perte de *welfare* est matérialisée par un triangle curviligne. Même lorsque le coût fixe du monopole est non nul, la surface de ce triangle curviligne incarne encore, précisément, la perte de *welfare* subie.

Nous venons de mettre en évidence le fait qu'un monopole (non contraint, non discriminant) prend une décision de production et de tarification qui, certes, lui sera favorable (il peut obtenir, à long terme, un profit microéconomique strictement positif), mais qui sera défavorable à la collectivité dans son ensemble. L'accroissement du surplus des producteurs ne compensera pas la baisse du surplus des consommateurs. Sur ce secteur d'activité, l'efficacité au sens de Pareto n'est plus atteinte. Ce résultat, obtenu à partir d'hypothèses très simples, est néanmoins très robuste : il est d'ailleurs un élément fondateur du droit de la concurrence. Le droit de la concurrence regroupe l'ensemble des règles et des principes qui doivent être respectés sur les différents marchés de biens ou services (sauf dérogation comme, par exemple, dans le secteur du spectacle sportif aux États-Unis). Les positions de monopole (ou, ce qui est de même nature, les cartels) sont jugées défavorables pour

la collectivité. Il existe donc une batterie d'instruments juridiques pour interdire ou sanctionner les positions dites « dominantes ». À l'inverse, le fonctionnement concurrentiel des marchés est favorisé, encouragé. Il existe une limite à ce principe : la présence d'un monopole naturel.

4 Le monopole naturel et son contrôle par la puissance publique

Nous avons, dès la première section, évoqué la possible existence d'une origine « économique » au monopole, c'est-à-dire une situation dans laquelle il serait pleinement justifié qu'une firme unique produise la totalité des unités du bien ou service. Ce qu'il faut entendre par là est que, du point de vue de l'efficacité productive, la présence de deux ou plusieurs firmes pour produire l'output serait inappropriée au sens où il serait globalement plus coûteux de faire produire, dans deux entités plutôt qu'une, toute quantité du bien ou service. Les caractéristiques de la technologie mobilisée pour produire sont un élément décisif. Si le coût marginal est croissant, il coûte de plus en plus cher de produire des unités supplémentaires ; on conçoit dès lors qu'il sera opportun « d'éclater » la production dans une multitude d'entités de production (confectionnant chacune un petit nombre d'unités d'output) plutôt que de tout produire dans une seule entité. À l'inverse, si le coût marginal était décroissant (le coût de production des unités supplémentaires est de plus en plus faible), il serait opportun de concentrer au maximum la production, et même de la concentrer en une seule et unique unité de fabrication. L'idée que le coût marginal puisse être décroissant est peu convaincante (cf. chapitre 6). En revanche, le coût unitaire peut l'être sur une échelle très étendue de production : il suffit en effet que le coût fixe de production soit très conséquent. Ainsi, tant que le coût unitaire sera décroissant, il sera opportun de produire toutes les unités d'output dans une seule et unique entité de production.

Récapitulons : si le coût marginal est décroissant (peu envisageable), le coût moyen le sera aussi et nous serons en présence d'un monopole naturel ; si le coût marginal est constant, le coût moyen est décroissant dès qu'il existe un coût fixe strictement positif et nous sommes aussi en présence d'un monopole naturel ; enfin, si le coût marginal est croissant, l'ampleur de l'éventuel coût fixe déterminera le seuil de production à partir duquel le coût moyen

deviendra croissant et ainsi l'échelle de production au-delà de laquelle il ne sera plus question d'un monopole naturel. Cette typologie est incomplète car le coût marginal n'est que rarement monotone (c'est-à-dire soit croissant, soit décroissant, soit constant sur tout son domaine, ici sur les valeurs positives de Q). Le coût marginal est souvent d'abord décroissant puis croissant, ce qui complique la description des circonstances dans lesquelles le coût unitaire s'avère être décroissant. Ce qui est décisif est en réalité **la position de la courbe de demande inverse au regard de celle de coût moyen**. Nous considérons ici qu'un monopole naturel existe tant que l'intersection entre la demande inverse et le coût moyen se trouve dans la partie décroissante du coût moyen.

Proposition

Lorsque la courbe de coût unitaire d'une firme est sécante avec la courbe de demande inverse dans la partie décroissante du coût unitaire, cette firme est en situation de **monopole naturel**. La décroissance du coût unitaire de production est une condition suffisante à l'existence d'un monopole naturel.

Nous représentons, sur la figure 6, les circonstances typiques de l'existence d'un monopole naturel : la courbe de demande inverse est sécante avec la courbe de coût moyen dans la partie décroissante du coût moyen.

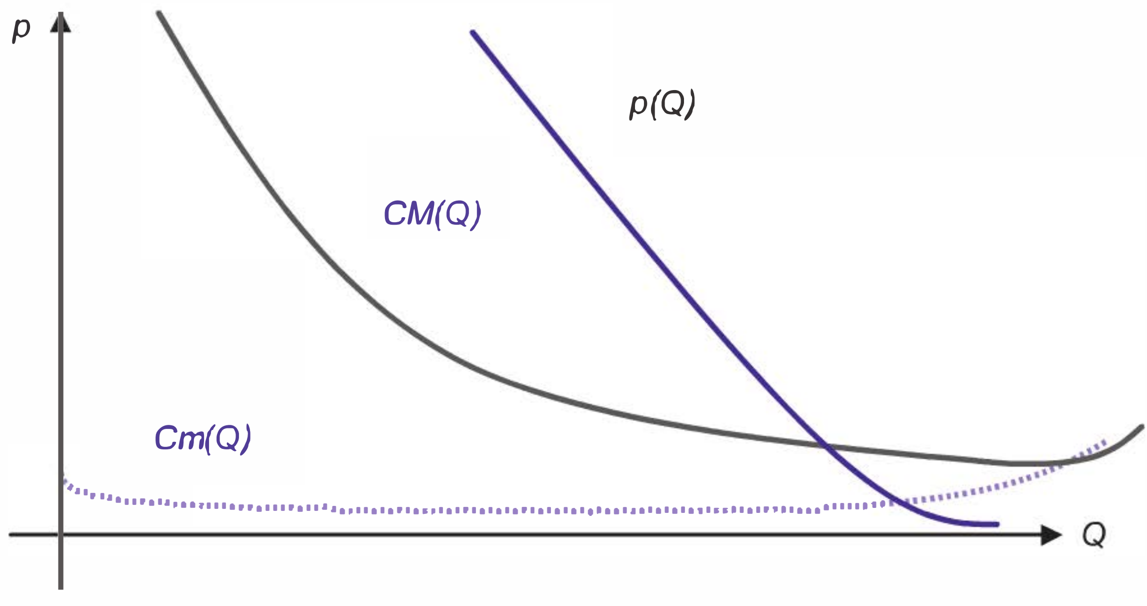


Figure 6 Monopole naturel

Nous l'avons dit dans la section 1, cette situation est très courante dans les secteurs d'activité reposant sur l'existence d'un réseau (de fils, de canalisations, de voies ferrées, de guichets, etc.). Dans le cas du transport ferroviaire, par exemple, si la partie de pur transport de voyageurs d'un point *A* à un point *B* est une activité pour laquelle l'efficacité productive ne commande pas qu'un unique opérateur soit présent, la partie de construction et d'entretien du réseau (voies, ponts et tunnels, signalisation, gares, etc.) est une activité pour laquelle, en revanche, il est économiquement « naturel » qu'une firme unique soit présente. Il en va de même pour la distribution d'électricité (construction et entretien des lignes aériennes ou enterrées, des transformateurs, des adductions aux particuliers) ou de gaz tandis que la production n'est pas un monopole naturel. Quant à la distribution d'eau, elle a une forte dimension locale (c'est un monopole naturel à l'échelle d'une aire urbaine) et au-delà de sa distribution, la captation de l'eau et son assainissement sont aussi des monopoles naturels.

Ces différentes activités d'entretien et de mise en valeur d'un réseau ou de distribution par un réseau sont généralement encadrées par la puissance publique au niveau local (eau potable) ou national (réseau ferré), soit par un contrôle direct (c'est une structure publique qui produit) soit par délégation (c'est une structure privée qui a reçu délégation de la puissance publique pour assurer la production). La raison d'être de cet encadrement public est évidente : si on laisse un entrepreneur administrer librement la structure en situation de monopole, il va exercer son pouvoir de marché, réaliser un profit microéconomique strictement positif et l'équilibre de long terme sur ce marché sera une situation non socialement optimale. C'est dans le but de restaurer l'efficacité économique, de relever le niveau de *welfare* globalement atteint, que la puissance publique contrôle le monopole. Concrètement, soit elle assure elle-même la production (en créant une régie ou une entreprise publique chargée de produire l'output), soit elle fixe les règles selon lesquelles un opérateur privé devra, par délégation, assurer la production de l'output.

Mais le statut de monopole naturel d'une activité n'est pas immuable. Les innovations technologiques ou l'accroissement de la demande peuvent être à l'origine de la perte du caractère « naturel » d'une situation monopolistique, pour tout ou partie de l'activité. Dans le cas d'une innovation, les courbes de coût vont s'abaisser : il devient moins coûteux de produire un

même nombre d'unités d'output. Il en résulte que l'intersection entre la demande inverse et la courbe de coût moyen, initialement dans la partie décroissante du coût moyen, peut désormais se situer dans la partie croissante du coût moyen, ce qui fait que le secteur d'activité perd son statut de monopole naturel (figure 7).

Sur la figure 8, un phénomène semblable se produit, mais en raison, cette fois-ci, de l'accroissement des débouchés, c'est-à-dire des quantités demandées pour tout niveau de prix.

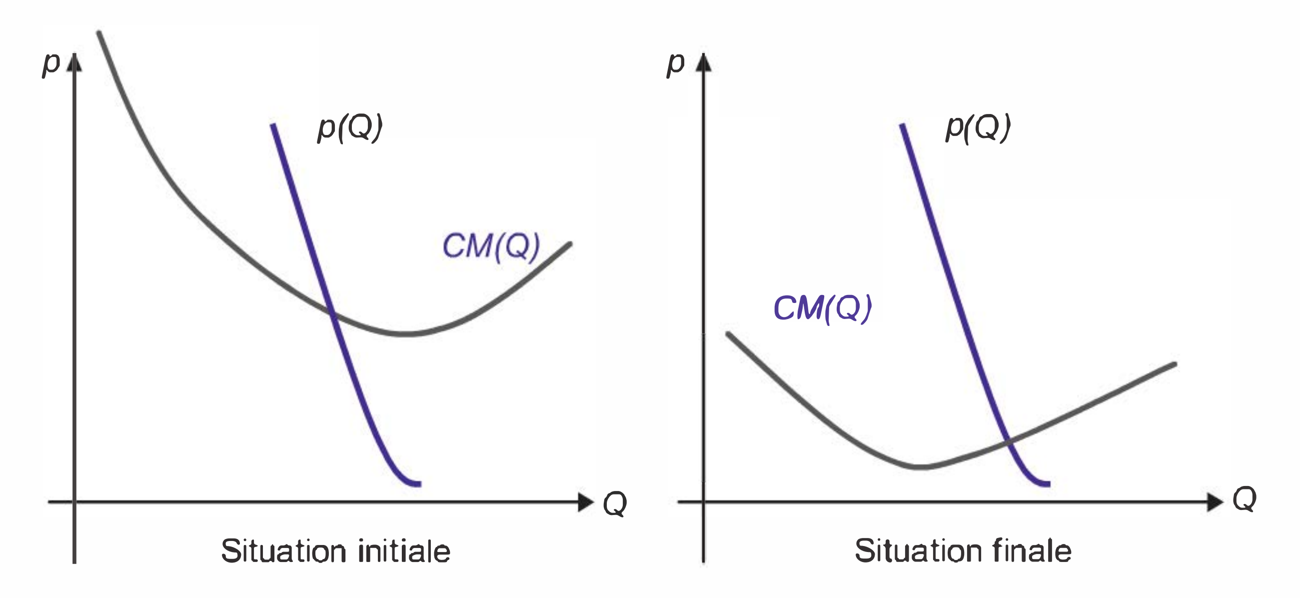


Figure 7 Perte du caractère naturel de la position monopolistique suite à une innovation technologique

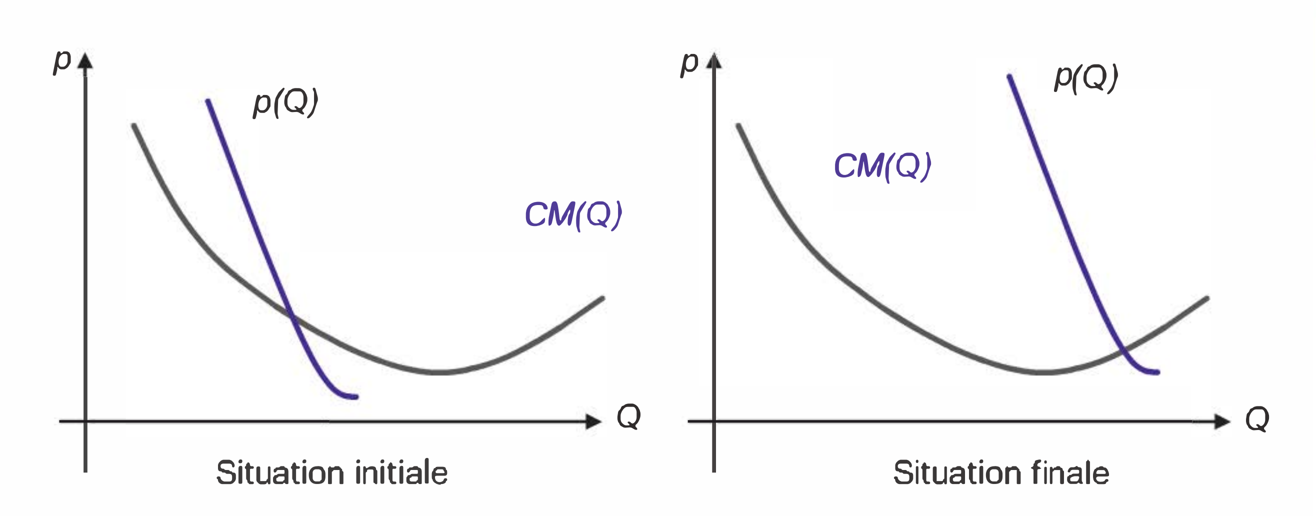


Figure 8 Perte du caractère naturel de la position monopolistique suite à un accroissement de la demande

Venons en maintenant au contrôle par la puissance publique des monopoles naturels. Deux possibilités principales s'offrent à la puissance publique. La première, qualifiée de solution de 1[er] rang (ou *first best*), consiste à choisir la tarification qui maximise, de manière inconditionnelle, le *welfare*. Cette tarification est une tarification au coût marginal. Mais, dès lors qu'existe un coût fixe strictement positif, cette tarification ne peut conduire le monopole qu'à des pertes (un profit microéconomique strictement négatif). C'est ce que nous pouvons observer sur la figure 9.

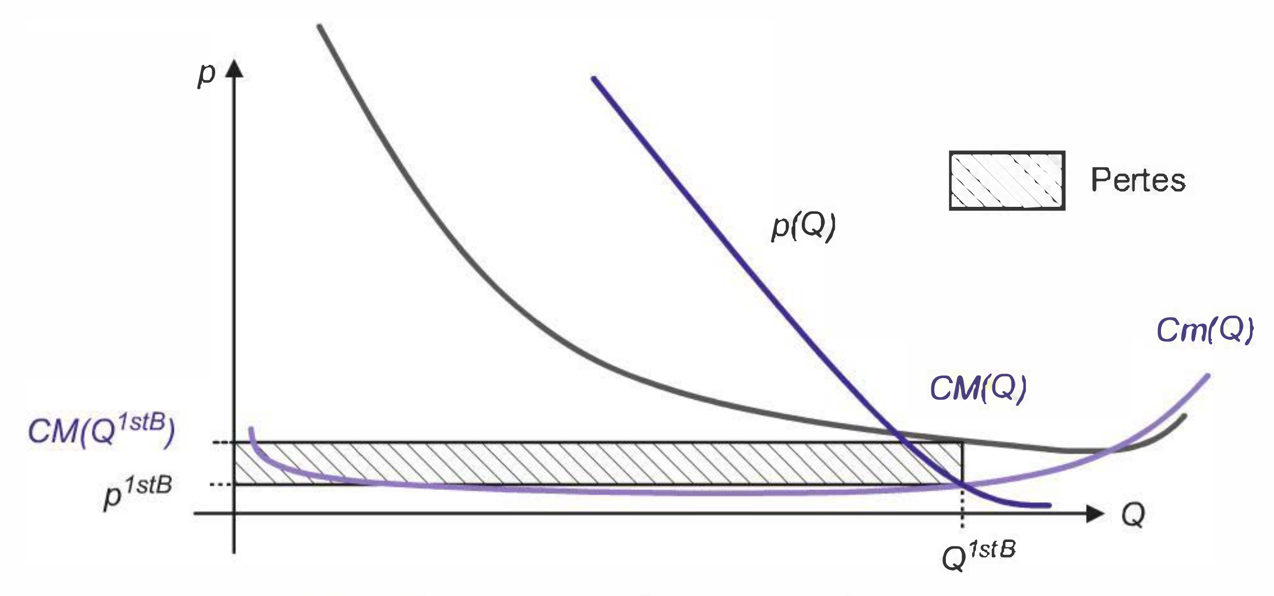


Figure 9 Tarification de 1[er] rang ou *first best*

En cas de tarification de 1[er] rang, notée p^{1stB}, le coût moyen relatif à la quantité produite Q^{1stB} sera nécessairement supérieur au prix, car, pour un monopole naturel, l'intersection entre le coût marginal et la demande inverse se fera nécessairement pour une quantité telle que le coût moyen est supérieur au coût marginal. Ainsi, $p^{1stB} < CM(Q^{1stB}) \Leftrightarrow p^{1stB} - CM(Q^{1stB}) < 0$ $\Leftrightarrow [p^{1stB} - CM(Q^{1stB})] \times Q^{1stB} < 0 \Leftrightarrow \Pi^{1stB} < 0$. Le monopole contraint de tarifer au coût marginal réalise donc des pertes. Ces pertes devront, d'une manière ou d'une autre, être supportées par la collectivité. En général, on met en place des transferts, c'est-à-dire des dispositifs reposant sur des impôts et taxes et dont le rôle est de subventionner les activités déficitaires. Il peut s'agir ici de déficits vertueux au sens où le déficit n'est pas la conséquence d'une mauvaise gestion de la structure productive, mais bien de la mise en place de règles de tarification visant à restaurer l'efficacité sociale.

Une **tarification de 1er rang** (*first best*) imposée par la puissance publique à une firme en situation de monopole est une tarification qui maximise, de manière inconditionnelle, le *welfare*.

La tarification de 1er rang est une tarification au coût marginal. Dans le cas d'un monopole naturel, cette tarification conduit à ce que la firme réalise un profit microéconomique négatif qui rend nécessaire la mise en place de transferts (pour financer le déficit).

Pour aller plus loin

Détermination de l'optimum de 1er rang

Comme nous l'avons indiqué plus haut, $W = SC + \Pi = \int_0^Q p(x)dx - C(Q)$.

La tarification de 1er rang est le résultat du programme de maximisation, sans contrainte, du *welfare* : $\underset{Q}{Max}\, W$

La condition d'optimalité est que $\frac{\partial W}{\partial Q} = 0 \Leftrightarrow p(Q) - \frac{\partial C(Q)}{\partial Q} = 0$ $\Leftrightarrow p(Q) = Cm(Q)$.

Bien que les pertes d'un monopole administré par la puissance publique puissent être le signe de la mise en œuvre d'une tarification propre à restaurer l'efficacité au sens de Pareto, les électeurs et les contribuables ne sont pas toujours disposés à accepter une telle situation. Il est donc opportun d'envisager une seconde possibilité. Cette seconde option, qualifiée de tarification de 2e rang (ou *second best*) consiste à choisir la tarification qui maximise le *welfare* sous condition que l'équilibre budgétaire du monopole soit atteint. En d'autres termes, il faut tenter d'obtenir le plus haut niveau d'efficacité sociale sous réserve que la tarification appliquée ne conduise pas le monopole à réaliser un profit négatif. La tarification retenue sera alors une tarification au coût moyen. Cette tarification est représentée sur la figure 10.

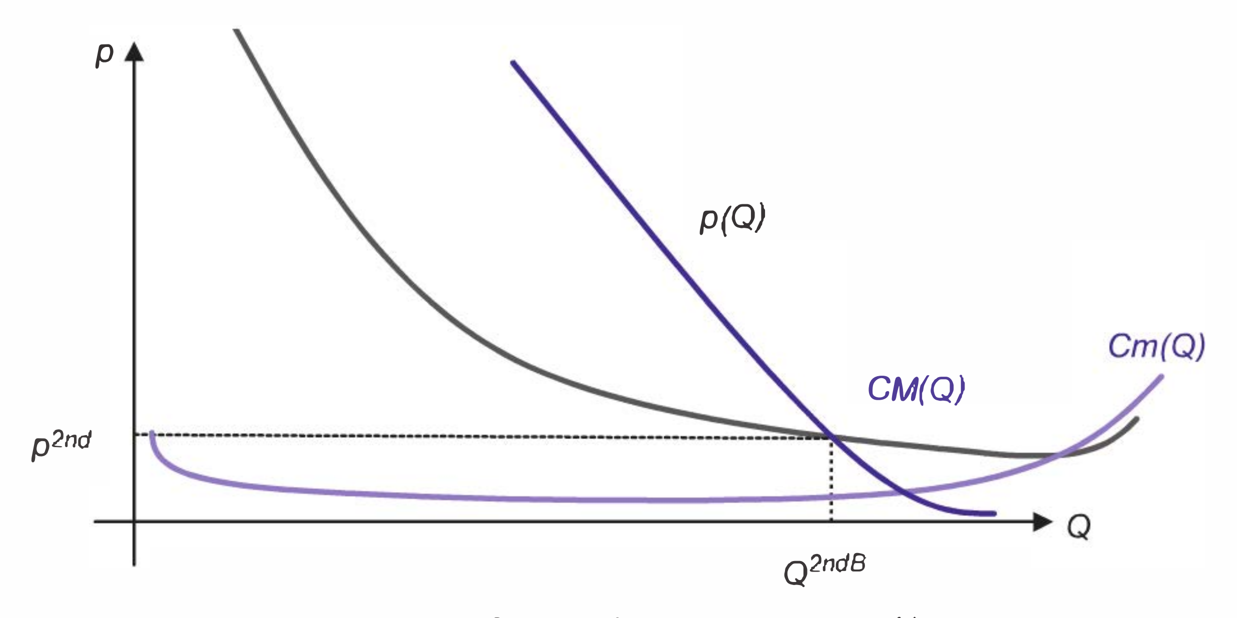


Figure 10 Tarification de 2e rang ou *second best*

En tarifant au coût moyen, la firme réalise un profit microéconomique nul. Au regard de la solution de 1er rang, le prix est plus élevé et la quantité produite plus faible. La tarification de 2e rang est d'autant plus élevée que la demande est faiblement élastique et que la contrainte d'équilibre budgétaire du monopole est pesante. Ceci est résumé par la règle de **Ramsey-Boiteux** qui dispose que :

$$\frac{p(Q)-Cm(Q)}{p(Q)}=\frac{\lambda}{(1+\lambda)|\varepsilon|}$$

où λ est un terme positif d'autant plus élevé qu'il est difficile de réaliser l'équilibre budgétaire du monopole.

Avec la tarification de 2e rang, le niveau de *welfare* finalement atteint est plus faible que celui atteint avec la tarification de 1er rang, mais sans qu'aucun déficit ne soit à combler, sans qu'aucun transfert n'ait à être mis en œuvre.

Une **tarification de 2e rang** (*second best*) imposée par la puissance publique à une firme en situation de monopole est une tarification qui maximise le *welfare* sous contrainte que le profit microéconomique de la firme soit non négatif.

La tarification de 2e rang est une tarification au coût moyen.

Pour aller plus loin

Règle de Ramsey-Boiteux

La tarification de 2e rang est le résultat du programme de maximisation suivant :

$$\underset{Q}{Max}W \quad s.c.\ \Pi \geq 0$$

On peut former un Lagrangien $\mathcal{L} = W + \lambda\Pi$ où λ désigne le multiplicateur de Lagrange associé à la contrainte de positivité du profit.

Les conditions d'optimalité sont que $\frac{\partial \mathcal{L}}{\partial Q} = 0$ et $\lambda\Pi = 0$.

Puisque $\mathcal{L} = \int_0^{\bullet} p(x)dx - C(Q) + \lambda[p(Q) \times Q - C(Q)]$

$$\frac{\partial \mathcal{L}}{\partial Q} = 0 \Leftrightarrow p(Q) - \frac{\partial C(Q)}{\partial Q} + \lambda\left[\frac{\partial p(Q)}{\partial Q} \times Q + p(Q) - \frac{\partial C(Q)}{\partial Q}\right] = 0$$

$$\Leftrightarrow [p(Q) - Cm(Q)](1+\lambda) = -\lambda p(Q)\left[\frac{\partial p(Q)}{\partial Q} \times \frac{Q}{p(Q)}\right]$$

$$\Leftrightarrow [p(Q) - Cm(Q)](1+\lambda) = -\lambda p(Q)\frac{1}{\varepsilon}$$

$$\Leftrightarrow \frac{p(Q) - Cm(Q)}{p(Q)} = \frac{\lambda}{(1+\lambda)|\varepsilon|}$$

La puissance publique peut également intégrer des **préoccupations d'équité** dans ses décisions de tarification en attribuant aux surplus des catégories de consommateurs défavorisés des « valeurs sociales » plus élevées qu'à ceux des catégories de consommateurs plus aisés. Ceci entraîne une modification de la règle de Ramsey-Boiteux visant à corriger à la baisse le prix des biens dont la demande est très inélastique mais dont la part est très importante dans le budget des consommateurs appartenant aux catégories sociales les plus défavorisées (tarif de l'électricité ou du gaz, par exemple).

Pour récapituler, nous montrons sur un même schéma (figure 11) les trois tarifications qui ont été évoquées dans cette section : la tarification non contrainte (p^M), la tarification de 1er rang (p^{1stB}) et la tarification de 2e rang (p^{2ndB}).

9

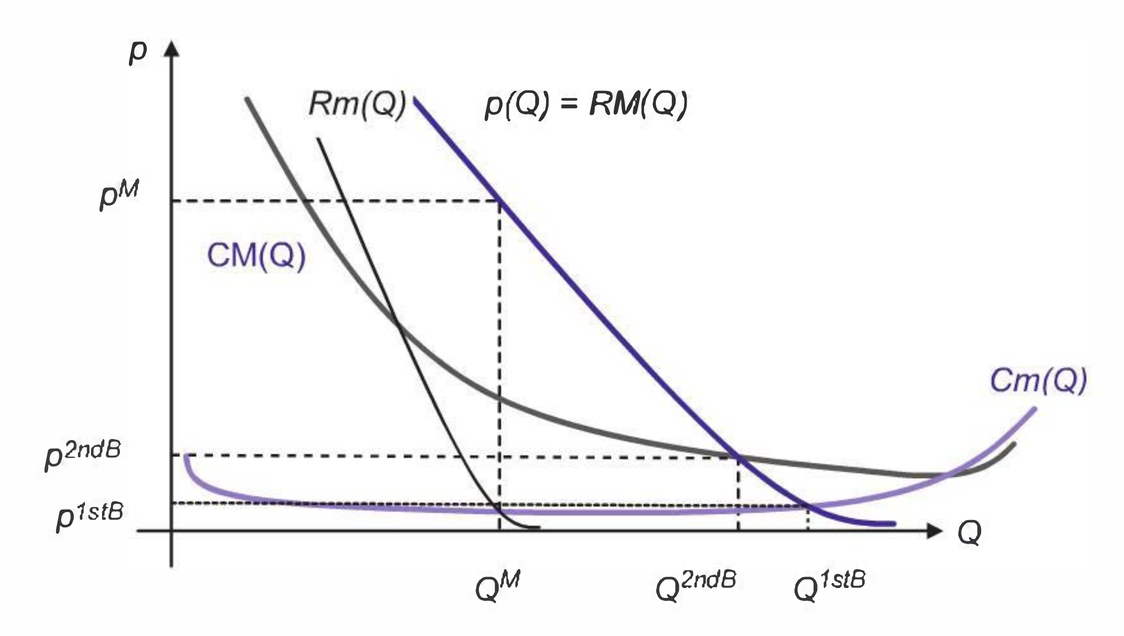


Figure 11 Tarifications non contrainte, de 1er rang et de 2e rang

5 La discrimination tarifaire

Jusqu'à présent, nous avons envisagé la décision d'un monopole non discriminant : chaque unité du bien ou service produit était vendue à un même prix unitaire aux différents consommateurs. Dans certaines circonstances, le monopole peut vendre les différentes unités produites à des prix différents.

- Un artiste peintre vend des toiles « comparables » à des prix différents à ses différents clients, selon la perception qu'il a de la disponibilité maximale à payer de chaque client. Dans cet exemple, qui correspond à une situation où la relation entre le producteur et le client est fortement individualisée, il nous est donné d'approcher le concept de **discrimination au 1er degré** ou **discrimination parfaite**. L'idée de discrimination au 1er degré correspond à une situation où les différentes unités du bien sont vendues à des prix tous différents. Plus explicitement, la firme serait, dans un tel cadre, en mesure de faire payer à chaque consommateur un prix unitaire égal à sa disponibilité maximale à payer.
- Un monopole local du transport urbain propose des tickets à l'unité ou des formules « 10 trajets », à des prix unitaires différents. Il s'agit

d'un cas de discrimination tarifaire fondée sur la quantité consommée du bien : un voyageur qui anticipe devoir utiliser souvent les transports publics de la ville achète des « carnets » de tickets, tandis qu'un voyageur occasionnel (un individu de passage dans la ville) achète des tickets à l'unité. Le prix unitaire du trajet est plus faible pour l'usager régulier que pour l'usager occasionnel. On parle dans ce cas de **discrimination tarifaire au 2nd degré**.

- Un monopole du transport ferroviaire vend un même trajet à un prix différent selon que le voyageur est un adulte ou en enfant (de moins de 12 ans). Il s'agit d'une discrimination liée à une caractéristique dite « exogène » du consommateur, telle que son âge, son sexe ou son lieu de résidence (le tarif d'accès à un équipement communal peut être différent selon que l'on réside dans la commune considérée ou dans une commune avoisinante). Une telle discrimination est qualifiée de **discrimination au 3e degré**.

La **discrimination au 1er degré** est une situation où une firme peut vendre son output aux différents consommateurs à des prix tous différents.

Les **discriminations au 2nd et 3e degrés** sont des situations où une firme peut vendre son output à des prix différents auprès de différentes catégories de consommateurs ; l'appartenance à une catégorie particulière de consommateurs est déterminée par la quantité consommée de l'output dans le cas de la discrimination au 2e degré, par une caractéristique exogène (âge, sexe, lieu de résidence, etc.) dans le cas de la discrimination au 3e degré.

La motivation principale d'un monopole qui pratique une discrimination tarifaire est de s'approprier tout ou partie du surplus des consommateurs en pratiquant des prix qui, schématiquement, s'approchent le plus possible de la disponibilité maximale à payer de chacun. De manière plus positive, une discrimination tarifaire permet aussi d'élargir la base des consommateurs susceptibles de jouir de la consommation d'un service ou de l'usage d'un équipement. Par exemple, si le tarif unique d'accès à une piscine

municipale est de 7 €, un certain nombre de familles modestes sont, de fait, exclues de son accès. Si l'on construit une tarification où chaque adulte paye 8 € et chaque enfant ne paye que 4 €, on peut, sans considérablement faire baisser les recettes de l'équipement, substantiellement accroître l'accessibilité (financière) de la piscine aux familles.

En plus des pratiques ci-dessus décrites, on observe l'existence de tarifications incluant une partie fixe. Par exemple, l'accès à la distribution d'eau potable, à la distribution d'électricité, à une connexion internet, etc. passe par l'acquittement d'une part forfaitaire ou « abonnement ». En sus de la part forfaitaire, on paye un montant variable dépendant du nombre de m^3 d'eau consommés ou, dans le cas des accès internet, une part variable généralement nulle. La raison d'être de l'existence d'une part fixe dans la tarification est la recherche d'un mode approprié de couverture du coût fixe de production. L'existence de monopoles naturels est souvent liée au niveau considérable du coût fixe dans le secteur considéré. Ceci entraîne des distorsions, au regard d'une situation concurrentielle, telles que celles que nous avons détaillées dans la section précédente. La tarification de 2e rang est une bonne réponse à ces distorsions mais on peut encore améliorer les choses : si, en plus de s'astreindre à la maximisation du *welfare* sous contrainte d'équilibre budgétaire du monopole, on met en place une tarification incluant une part fixe, il est possible d'obtenir un niveau de *welfare* encore plus élevé. Puisque l'origine de la distorsion est l'existence de la part fixe du coût, la correction de la distorsion peut être obtenue en faisant supporter le coût fixe au consommateur sous la forme d'une part elle-même fixe du tarif. C'est comme si chacun, lorsqu'il paye sa facture d'eau, payait, au travers de l'abonnement, un petit morceau du réseau qui permet à l'eau potable d'arriver dans son logement.

Ce qui complique parfois le décryptage des tarifications est le fait que sont parfois mélangées tarification incluant une partie fixe et discrimination tarifaire (au 2e ou 3e degré). Par exemple, les « grands » voyageurs sur un réseau ferré peuvent acheter un abonnement qui leur permet de payer chacun des trajets à moitié prix (2e degré) ; les seniors, peuvent, eux aussi choisir d'acheter un abonnement (qui leur est spécifiquement dédié) et grâce auquel ils ne payent qu'une fraction du plein tarif (3e degré).

La discrimination tarifaire est une pratique très répandue et qui ne concerne pas seulement les monopoles. Elle permet, aux firmes qui la pratiquent, de s'approprier une partie du surplus des consommateurs. Elle est d'autant plus facile à mettre en place que la relation entre la firme et le consommateur est individualisée. Le consommateur ne sort, *a priori*, pas gagnant des pratiques de discrimination tarifaire. À une époque de ciblage performant des caractéristiques du consommateur et d'individualisation accrue de la relation client, ce dernier a du souci à se faire...

Pour conclure ce chapitre, nous mentionnerons l'existence de situations où une firme, plutôt qu'être en situation de monopole de l'offre, est un monopole de la demande. On qualifie cette situation de monospone.

Une firme est en position de **monopsone** si elle est l'unique acheteur d'un de ses facteurs de production.

Comme pour un monopole, il s'agit d'une position de force qui pourra être exploitée par la firme, au détriment, cette fois, de ses fournisseurs. Comme pour le monopole, il est désormais envisageable que la firme parvienne à obtenir un profit microéconomique strictement positif à long terme, non pas, ici, en raison de sa capacité à vendre l'output à un prix supérieur à son coût marginal, mais en raison de sa capacité à payer un des facteurs de production moins cher que le prix concurrentiel.

10 L'oligopole

Mots-clés

Interactions stratégiques, théorie des jeux, fonction de meilleure réponse, équilibre non coopératif, duopole de Cournot, duopole de Stackelberg, duopole de Bertrand, collusion.

Dans les chapitres 8 et 9, deux cas polaires de la décision d'offre de la firme ont été étudiés : le cas d'une firme en situation de concurrence pure et parfaite et le cas d'un monopole. Ces deux cas présentent, pour des raisons diamétralement opposées, un trait commun : il n'est pas besoin d'envisager l'existence d'interactions stratégiques entre les firmes. En d'autres termes, lorsqu'elle prend sa décision, la firme n'a pas besoin de tenir compte de la décision (ou de ce qu'elle anticipe devoir être la décision) d'une quelconque firme concurrente. En situation de concurrence pure et parfaite, les firmes n'éprouvent pas le « besoin » de se préoccuper des décisions de leurs voisines car le prix de vente de l'output s'impose à elles ; à l'opposé, le monopole n'éprouve pas plus le besoin de se préoccuper de décisions de concurrents… car il n'en a pas ! Dans ces deux cas extrêmes, la rivalité entre les firmes n'est pas internalisée dans l'analyse. Le présent chapitre est consacré à toutes les configurations de marché intermédiaires, où il n'est plus possible d'ignorer que la décision de chaque firme est prise en tenant compte du comportement de ses rivales et, mieux encore, des réactions de ses rivales à ses propres décisions.

La discipline qui permet d'étudier ces situations est la théorie des jeux, branche des mathématiques qui étudie les décisions des individus (ou des firmes, ou des nations, ou des … joueurs d'échec) lorsque ceux-ci tiennent

compte, pour prendre leur propre décision, des décisions de leurs adversaires. Née au milieu du XX[e] siècle, elle est utile aux stratèges militaires comme aux économistes qui s'intéressent au comportement des firmes dans un secteur d'activité donné. Elle repose sur la compréhension de ce que des décideurs avertis, lorsqu'ils font leurs choix, tiennent compte des décisions de leurs adversaires et n'ignorent pas que leurs adversaires font de même. Ainsi, chacun anticipe les réactions de ses concurrents, mais aussi les réactions des concurrents aux réponses que le joueur apporte à la réaction des concurrents à ses propres actions, et ainsi de suite, comme raisonne un joueur d'échec qui réfléchit à la prochaine séquence de coups entre lui et son adversaire, ainsi qu'à la séquence suivante, ainsi qu'à celle d'après, etc.

1 Quelques notions de théorie des jeux

La théorie des jeux est une discipline mathématique consacrée à l'étude des interactions stratégiques entre les agents. Une situation est « stratégique » lorsque la satisfaction éprouvée par un « joueur » (un participant au « jeu ») dépend non seulement de son choix, mais aussi des choix effectués par un ou plusieurs autres individus. En théorie des jeux, la « satisfaction » éprouvée à la suite d'une décision est entendue au sens large : il peut s'agir d'un gain ou d'une perte monétaire, d'un gain ou d'une perte en nature, de satisfaction ou d'insatisfaction morale ou physique, d'une amélioration ou d'un dommage à la santé... Pour désigner la satisfaction éprouvée par les joueurs, on utilise indifféremment les termes satisfaction, utilité, gain ou paiement. À l'aide d'un exemple simple, nous allons tenter de comprendre comment il est possible d'analyser les « jeux » et comment il est possible de parvenir à un éventuel équilibre non coopératif.

L'exemple choisi est connu sous le nom de « jeu de l'urne » : on demande à deux joueurs de mettre dans une urne opaque soit un billet de 20 €, soit « rien du tout ». Chacun des deux joueurs ne voit pas ce que l'autre fait. À la fin du jeu, le contenu de l'urne est multiplié par 1,5 puis redistribué en parts égales aux 2 joueurs. Pour résumer le jeu (et pour aider à imaginer les ques-

tions que se posent les participants), on le représente généralement sous forme d'un tableau où figurent les paiements nets au bénéfice des joueurs (c'est-à-dire ce qu'ils obtiennent à la fin du jeu, en ayant soustrait les « mises » initiales). C'est ce qui figure dans le tableau 1 où les stratégies (les options possibles) du joueur 1 apparaissent en ligne (en noir) et les stratégies du joueur 2 apparaissent en colonne (en bleu). Dans les cases du tableau apparaissent des couples de paiements : la première valeur est le paiement obtenu par le joueur 1, la seconde valeur est le paiement obtenu par le joueur 2.

Tableau 1 Matrice de paiements du « jeu de l'urne »

Joueur 1 \ Joueur 2	Met 0 €	Met 20 €
Met 0 €	(0 ; 0)	(15 ; –5)
Met 20 €	(–5 ; 15)	(10 ; 10)

Si un joueur met 20 € dans l'urne et que son adversaire ne met rien, le montant trouvé dans l'urne est de 20 €. En multipliant cette somme par 1,5, on obtient le montant qui sera redistribué à parts égales entre les deux joueurs soient 20 × 1,5 = 30 €. Chacun recevra donc 15 €, ce qui correspond à un gain net de 15 € pour celui qui n'avait rien mis et à une perte nette de 5 € pour celui qui avait mis 20 €. En envisageant ainsi toutes les possibilités, on obtient les paiements qui apparaissent dans le tableau 1. Cet exemple simple permet de percevoir les « sentiments » contradictoires qui assaillent les participants à un tel jeu : chacun a individuellement intérêt à ne rien mettre dans l'urne (en espérant que l'autre ne fera pas de même) mais les deux joueurs auraient collectivement intérêt à mettre chacun 20 € dans l'urne.

L'équilibre non coopératif d'un tel jeu, s'il existe, est l'intersection des *meilleures réponses* des joueurs. La meilleure réponse d'un joueur à une quelconque stratégie de son concurrent est la stratégie pour laquelle il obtiendrait le plus fort paiement. Pour identifier les meilleures réponses des deux joueurs, on fait figurer un astérisque (*) au-dessus de chacun des paiements les plus élevés obtenus par chaque joueur « étant donnée » la stratégie de l'autre. Le (les) équilibre(s) non coopératif(s), s'il(s) existe(nt),

corresponde(nt) au(x) couple(s) de stratégies pour le(s)quel(s) le couple de paiements sera signalé par deux astérisques (**).

Identifions, dans le tableau 2, les meilleures réponses de chacun des deux joueurs aux stratégies potentiellement choisies par le concurrent. Donnons l'exemple de la détermination, par le joueur 1, de ses meilleures réponses : si le joueur 2 venait à choisir la stratégie « Mettre 0 € », la meilleure réponse du joueur 1 serait « Mettre 0 € » qui conduit à un paiement (0 €) plus élevé que le paiement obtenu en optant pour la stratégie « Mettre 20 € » (paiement qui serait de – 5 €). De même, si le joueur 2 venait à choisir la stratégie « Mettre 20 € », la meilleure réponse du joueur 1 serait « Mettre 0 € » qui conduit à un paiement (15 €) plus élevé que celui qu'il obtiendrait en retenant la stratégie « Mettre 20 € » (paiement de 10 €). La détermination, par le joueur 2, de ses meilleures réponses, se fait par un raisonnement semblable (et est, dans le cas présent, parfaitement symétrique).

Tableau 2 Meilleures réponses des joueurs aux stratégies de l'adversaire

Joueur 2 / Joueur 1	Met 0 €	Met 20 €
Met 0 €	* * (0 ; 0)	* (15 ; –5)
Met 20 €	* (–5 ; 15)	(10 ; 10)

On observe l'existence d'un seul et unique équilibre non coopératif : c'est le couple stratégique (Met 0 € ; Met 0 €) qui conduit au couple de paiements (0 ; 0). Cet équilibre non coopératif incarne à merveille les situations où l'absence de coopération conduit chacun à constater *a posteriori*, avec dépit, qu'il obtient moins que ce qu'il aurait obtenu si un comportement coopératif conjointement respecté avait été décidé. Il n'est pas ici question de disserter sur ce que serait la société si l'ensemble de ses membres adoptait, de manière spontanée et systématique, un comportement coopératif. Notre seul objectif est d'indiquer que, dans un contexte où les participants au jeu sont de purs rivaux, il est pertinent de chercher à

identifier le ou les équilibre(s) non coopératif(s) s'il(s) existe(nt). De plus, il faut avoir conscience que, dans une optique de pure rivalité, même si les joueurs s'entendent au préalable sur un couple stratégique s'apparentant à une coopération – ici, le couple (Mettre 20 € ; Mettre 20 €) – chacun a intérêt à dévier unilatéralement de son engagement (cf. section 5). En outre, lorsque les firmes font effectivement le choix d'une entente, même si les paiements (en l'occurrence les profits) des 2 firmes qui produisent l'output sont conjointement accrus, la perte de surplus essuyée par les consommateurs étant supérieure au gain obtenu par les producteurs, le *welfare* de la société dans son ensemble est dégradé. En conséquence, la coopération entre les firmes n'a ici rien de souhaitable. C'est ce que le droit de la concurrence affirme, en substance, dans ses grands principes.

On appelle **équilibre non coopératif** d'un jeu une combinaison stratégique telle que chaque stratégie est une meilleure réponse aux stratégies des autres joueurs.

L'équilibre non coopératif est aussi appelé équilibre de Nash ou équilibre de Cournot-Nash.

Nous avons envisagé ci-dessus un jeu pour lequel il n'y a qu'un nombre fini de stratégies pour chaque joueur (deux en l'occurrence). Supposons désormais que la décision ayant un caractère stratégique prenne la forme, pour chaque joueur, du choix d'une valeur numérique réelle x (prix, quantité d'output, montant de dépenses publicitaires, etc.) comprise entre 0 et $+\infty$. Précisément, la décision du joueur 1 sera une valeur x_1 et la décision du joueur 2 sera une valeur x_2 : il s'agit maintenant d'envisager un nombre infini non dénombrable de stratégies pour chaque joueur. Dans ce type de jeux, les paiements obtenus par les joueurs deviennent des *fonctions de paiements* (et non plus des valeurs numériques ponctuelles comme auparavant). Choisissons de désigner par $\Pi_1(x_1 ; x_2)$ et $\Pi_2(x_1 ; x_2)$ ces fonctions de paiement. Comme précédemment, l'équilibre non coopératif du jeu est défini comme une combinaison stratégique pour laquelle la stratégie de chaque joueur est une meilleure réponse à la stratégie de l'autre joueur. Dans le contexte des jeux avec fonctions de paiements, la meilleure réponse

prend donc la forme d'une *fonction de meilleure réponse*. Les fonctions de meilleures réponses des joueurs 1 et 2 sont respectivement notées MR_1 (x_2) et MR_2 (x_1). L'équilibre non coopératif sera le couple stratégique $(x_1^* ; x_2^*)$ vérifiant :

$$\begin{cases} x_1^* = MR_1(x_2^*) \\ x_2^* = MR_2(x_1^*) \end{cases}$$

En d'autres termes, l'équilibre non coopératif du jeu est l'intersection des fonctions de meilleures réponses.

Pour aller plus loin

Comment calculer les fonctions de meilleures réponses ?

Pour calculer la fonction de meilleure réponse d'un joueur, il faut maximiser la fonction de paiement de ce joueur relativement à la variable sur laquelle porte sa décision. Ainsi, pour obtenir $x_1 = MR_1$ (x_2), il faut $\underset{x_1}{Max}\,\Pi_1$ $(x_1 ; x_2)$ et pour obtenir la relation $x_2 = MR_2$ (x_1), il faut $\underset{x_2}{Max}\,\Pi_2$ $(x_1 ; x_2)$. Les fonctions obtenues, MR_1 (x_2) et MR_2 (x_1) peuvent alors être introduites dans un système de 2 équations à 2 inconnues dont la solution est l'intersection des meilleures réponses, autrement dit le couple stratégique $(x_1^* ; x_2^*)$ qui constitue l'équilibre non coopératif du jeu.

2 Le duopole de Cournot

Si la notion d'équilibre non coopératif est due au mathématicien américain John Nash, son intuition initiale est apparue un siècle plus tôt, en 1838, dans les travaux d'un ingénieur économiste français : Antoine-Augustin Cournot. Cournot a étudié un marché où seules deux firmes produisent un même bien ou service et le proposent aux consommateurs. Dans sa construction, Cournot considère deux firmes de forces égales, c'est-à-dire telles qu'aucune ne possède d'ascendant sur l'autre. On peut imaginer deux entreprises de tailles à peu près semblables, arrivées en même temps sur le marché et produisant des biens très peu différenciés, par exemple deux constructeurs aéronautiques fabricant des

gros porteurs ou deux producteurs de semences agricoles (à l'échelle mondiale). En dépit de son manque de réalisme (les firmes ne produisent jamais des bien indifférenciés), cette hypothèse a la grande vertu de permettre un traitement très simple du jeu auquel participent les firmes. En retenant l'hypothèse d'homogénéité du bien, il ne pourra exister qu'une valeur unique du prix du bien (ce sans quoi, toute la demande s'adresserait à la firme pratiquant le prix le plus bas, alors que sa concurrente aurait vu s'enfuir tous les consommateurs). C'est désormais sur la quantité produite du bien ou service que va porter la décision centrale. En d'autres termes, la variable stratégique, dans le modèle Cournot, est la quantité.

On peut critiquer ce modèle au prétexte que, dans la réalité, ce sont les prix qui sont les principales variables stratégiques pour les firmes. C'est en substance ce que fit (de manière excessivement véhémente) Joseph Bertrand en 1883 (cf. section 4). Mais on peut aussi voir dans le choix de Cournot une option astucieuse pour figurer, par un chemin détourné, les caractéristiques de l'équilibre non coopératif d'un marché que se partagent deux firmes : certes, dans la réalité, la première (ou la principale) variable stratégique n'est pas la quantité, mais les mécanismes mis en évidence par Cournot sont très proches des mécanismes existants, et représentatifs de la démarche d'optimisation des entreprises, et de l'équilibre auquel celles-ci parviennent.

On appelle **oligopole** un secteur d'activité pour lequel un petit nombre de firmes est en concurrence. Les firmes constituant l'oligopole s'identifient parfaitement les unes les autres. Elles prennent des décisions en tenant compte des décisions de leurs concurrentes ainsi que des réactions de leurs concurrentes à leurs propres décisions et ainsi de suite.

Lorsque l'oligopole n'est constitué que de deux firmes, on parle d'un **duopole**.

Le **duopole de Cournot** est un duopole symétrique au sens où aucune firme ne possède d'ascendant d'aucune sorte sur sa concurrente. Cela ne signifie pas que les firmes aient nécessairement les mêmes fonctions de

coût : celle qui a la meilleure technologie réalisera évidemment un profit supérieur à sa concurrente, mais le fait d'avoir des coûts plus élevés ne remet en cause, ni la présence de la seconde firme, ni le postulat qu'elle se trouve sur un pied d'égalité avec sa performante rivale. Il y a certainement une limite à la disparité des coûts des deux firmes pour que l'on puisse concevoir qu'elles interagissent effectivement sans qu'existe un ascendant de l'une sur l'autre (ce qui est la marque du cadre hypothétique du duopole de Cournot) : si l'une des deux supportait des coûts considérablement plus élevés que sa rivale, peut-être obtiendrait-elle un profit microéconomique négatif à l'équilibre non coopératif du jeu. Pourrait-elle alors demeurer sur le marché ? Ou devrait-elle reconsidérer sa position face à son puissant rival ? En filigrane de cette discussion, nous pointons la fragilité de la notion de configuration « symétrique » sans référence aux caractéristiques de coûts des firmes. Pour éviter cet écueil, nous étudions l'équilibre non coopératif du jeu de Cournot en supposant que les fonctions de coûts des deux firmes sont semblables.

Hypothèses du duopole de Cournot

Le duopole de Cournot est une configuration de marché dans laquelle deux firmes en positions symétriques se font concurrence en quantité. Elles produisent et vendent un bien homogène à un prix unique p. Chacune connaît les caractéristiques de la relation entre prix et quantité demandée, les caractéristiques de sa propre fonction de coût ainsi que les caractéristiques de la fonction de coût de sa concurrente.

Cette dernière précision sur la parfaite connaissance de la fonction de coût de la concurrente fait référence au cadre d'information parfaite dans lequel nous nous situons. Il s'agit, là d'une hypothèse peu réaliste[1].

La demande inverse est supposée être, comme précédemment, une fonction décroissante $p(Q)$. Les deux joueurs (les deux firmes) sont indicées 1 et 2. La firme 1 décide de sa production Q_1, la firme 2 décide de sa produc-

1. Dans le chapitre 12, nous verrons comment il est possible, en microéconomie, de dépasser le cadre de l'information parfaite.

tion Q_2. La quantité totale produite Q est la somme simple des quantités produites par les firmes 1 et 2 : $Q = Q_1 + Q_2$. La fonction de coût total de la firme 1 est $C_1(Q_1)$. La fonction de coût total de la firme 2 est $C_2(Q_2)$. L'équilibre non coopératif du jeu de Cournot sera la combinaison stratégique $(Q_1^* ; Q_2^*)$ telle que :

$$\begin{cases} Q_1^* = MR_1(Q_2^*) \\ Q_2^* = MR_2(Q_1^*) \end{cases}$$

L'**équilibre non coopératif du jeu de Cournot** est la combinaison stratégique $(Q_1^* ; Q_2^*)$ telle que la quantité produite par chaque firme est une meilleure réponse à la quantité produite par sa concurrente.

On peut représenter cet équilibre non coopératif en traçant les fonctions de meilleures réponses de chacune des deux firmes dans un repère où la quantité produite par la firme 1 figure en abscisse et la quantité produite par la firme 2 en ordonnée. Au point C, à l'intersection de ces fonctions de meilleures réponses, figure l'équilibre non coopératif du jeu. Cet équilibre non coopératif apparaît sur la figure 1.

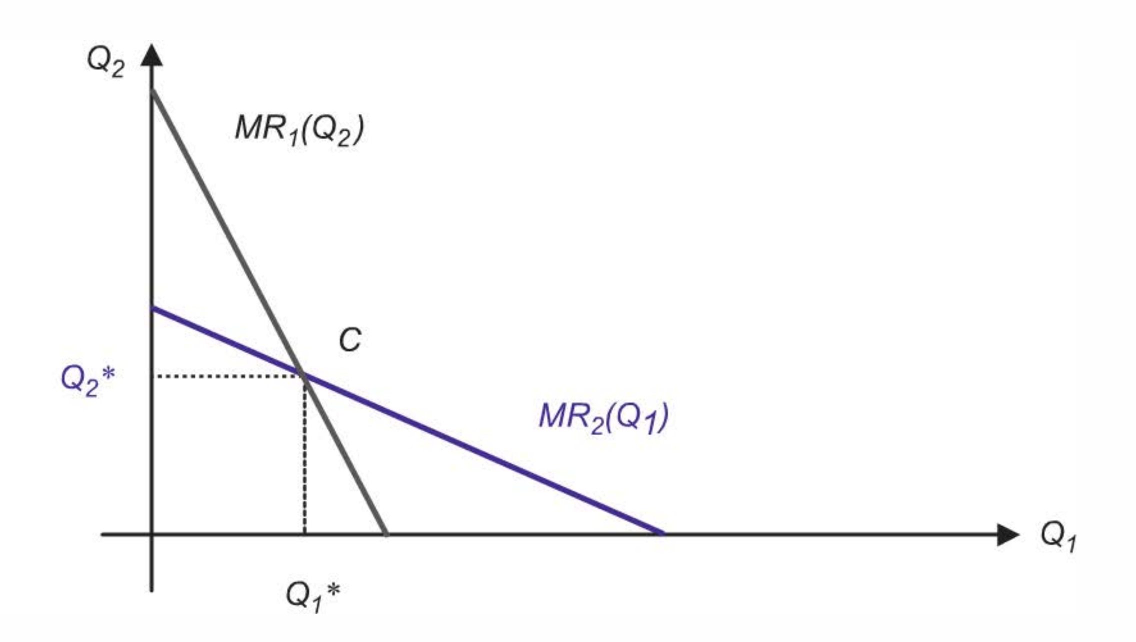


Figure 1 Équilibre non coopératif du duopole de Cournot

Pour aller plus loin

Détermination de l'équilibre de Cournot

Pour déterminer sa fonction de meilleure réponse, la firme 1 doit maximiser son profit (sa fonction de « paiement ») relativement à la variable sur laquelle porte sa décision, c'est-à-dire la quantité Q_1. Ainsi, elle $\underset{Q_1}{Max}\ \Pi_1\ (Q_1\,;Q_2) = p(Q_1 + Q_2) \times Q_1 - C_1(Q_1)$. La condition d'optimalité est que $\frac{\partial \Pi_1(Q_1;Q_2)}{\partial Q_1} = 0 \quad \Leftrightarrow Q_1 = MR_1\ (Q_2)$. De même la firme 2 doit maximiser son profit relativement à la variable Q_2.

Elle $\underset{Q_2}{Max}\ \Pi_2\ (Q_1\,;Q_2) = p(Q_1 + Q_2) \times Q_2 - C_2(Q_2)$.

La condition d'optimalité est que $\frac{\partial \Pi_2(Q_1;Q_2)}{\partial Q_2} = 0 \quad \Leftrightarrow Q_2 = MR_2\ (Q_1)$.

À l'équilibre non coopératif de Cournot, les firmes peuvent réaliser un profit microéconomique durablement positif. On peut montrer qu'à technologie donnée, le profit total obtenu par les deux firmes est (sous certaines hypothèses) inférieur au profit qu'aurait obtenu un monopole. En cela, le duopole est bien une situation intermédiaire entre monopole et concurrence pure et parfaite.

Si l'on prolonge l'analyse, on peut, sous ces mêmes hypothèses, montrer que le profit total qu'obtiendraient trois firmes dans un « triopole » de Cournot serait inférieur au profit total obtenu par les deux firmes d'un duopole et ainsi de suite. Lorsque le nombre de firmes qui forment l'oligopole devient très grand, le profit total tend vers zéro et le prix pratiqué par les firmes à l'équilibre non coopératif de l'oligopole tend vers le prix de concurrence pure et parfaite. Dans cette optique, la configuration de concurrence pure et parfaite s'apparente à un jeu de Cournot dans lequel le nombre de joueurs tendrait vers l'infini.

Le fait que les firmes réalisent un profit microéconomique strictement positif dans le cadre d'un duopole de Cournot est le signe qu'un tel équilibre n'est pas une configuration socialement efficace : le prix pratiqué par les firmes est plus élevé que le prix qui serait pratiqué sur un marché de concurrence pure et parfaite et la quantité totale produite est plus faible que ce qui serait décidé dans un environnement concurrentiel. Ce caractère non Pareto-optimal de l'équilibre non coopératif de Cournot peut aussi être mis en évidence en éta-

blissant que les deux firmes pourraient, dans le cadre d'un comportement coopératif, accroître conjointement leur « satisfaction » au regard de ce qu'elles obtiennent dans un cadre non coopératif. Pour le montrer, nous pouvons nous appuyer sur la représentation de leurs courbes d'iso-profit.

Comme son nom l'indique, la courbe d'iso-profit d'une firme désigne toutes les combinaisons de quantités produites qui conduisent à l'obtention d'un même et unique niveau de profit. Ces courbes ne sont pas très simples à tracer. Elles respectent néanmoins une propriété qui semble évidente : toute courbe d'iso-profit d'une firme passe par un maximum là où elle est sécante avec la fonction de meilleure réponse de la firme considérée. Représentons, sur la figure 2, quelques courbes d'iso-profit de la firme 1.

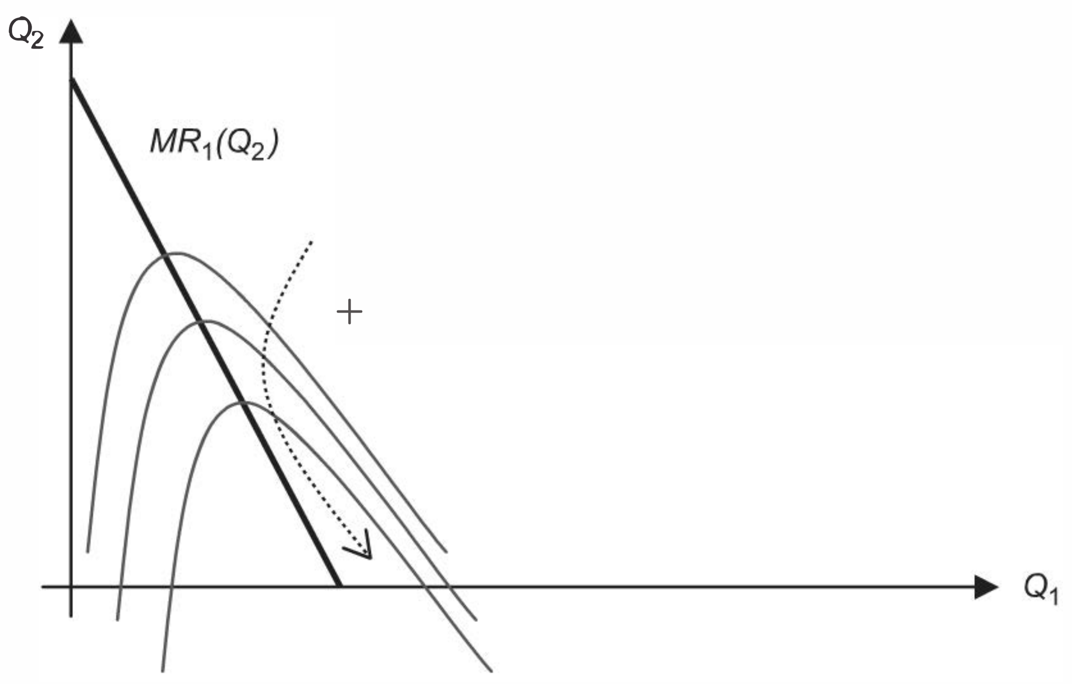


Figure 2 Courbes d'iso-profit de la firme 1

Il y a autant de courbes d'iso-profit que de valeurs possibles du profit. On peut établir que le niveau Π_1 $(Q_1 ; Q_2)$ de profit obtenu par la firme 1 est d'autant plus élevé que la courbe se situe vers le bas et vers la droite.

On peut également représenter les courbes d'iso-profit de la firme 2. Dans le cas où les deux firmes ont les mêmes fonctions de coût, leurs courbes d'iso-profit sont symétriques relativement à la première bissectrice. De manière générale, on peut montrer que le niveau Π_2 $(Q_1 ; Q_2)$ de profit atteint est d'autant plus élevé que la courbe se situe vers le haut et vers la gauche. De telles courbes apparaissent sur la figure 3.

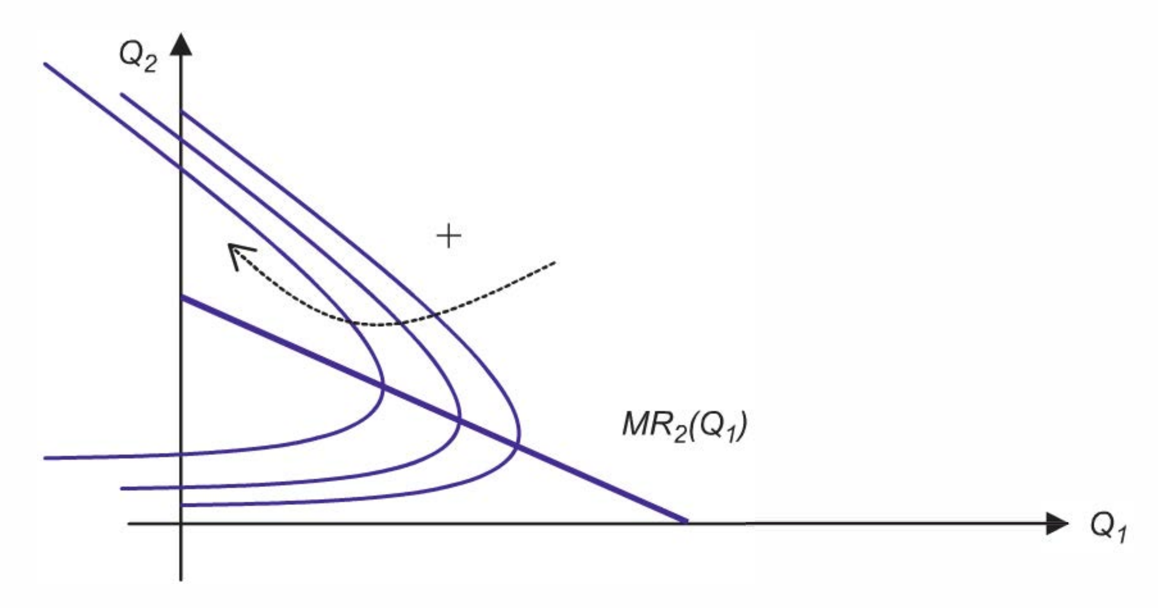


Figure 3 Courbes d'iso-profit de la firme 2

Si l'on figure maintenant les courbes d'iso-profit des firmes 1 et 2 passant par l'équilibre non coopératif du jeu de Cournot, on observe la présence d'une lentille dont l'un des sommets est le point C, c'est-à-dire le couple (Q_1* ; Q_2*) d'équilibre. La présence d'une telle lentille est, comme pour la situation étudiée au sein de la boîte de Pareto-Edgeworth dans le chapitre 5, le signe que la combinaison stratégique (Q_1* ; Q_2*) n'est pas un optimum de Pareto (figure 4).

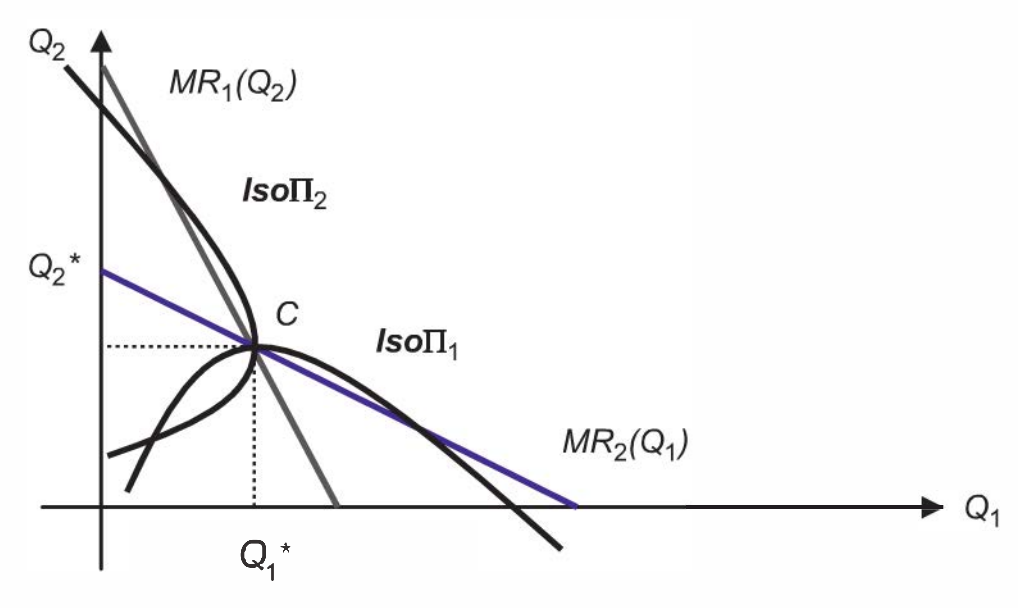


Figure 4 L'équilibre non coopératif du duopole de Cournot n'est pas un optimum de Pareto

L'existence d'une lentille, ou zone d'avantages mutuels, matérialise l'existence d'un ensemble de couples de quantités produites qui permettraient aux deux firmes d'obtenir, conjointement, un profit plus élevé. L'équilibre non coopératif C n'est donc pas un optimum de Pareto. L'inefficacité sociale du duopole de Cournot signifie à la fois que les firmes gagneraient à coopérer plutôt qu'à se comporter en pures rivales, mais aussi que les consommateurs gagneraient à ce qu'il y ait plus de firmes présentes sur le marché (et que celles-ci ne coopèrent pas). Sauf si la configuration « naturelle » du marché est un duopole. Dans le chapitre précédent, nous avons indiqué que les conditions technologiques pouvaient conduire à ce que la configuration de marché conduisant à l'efficacité sociale soit la configuration monopolistique (cas du monopole naturel). Certes, l'obtention de l'efficacité sociale nécessitait l'intervention de la puissance publique, mais l'efficacité productive était susceptible d'être atteinte avec une firme en situation de monopole. On peut imaginer que, dans certaines conditions technologiques, la configuration de marché susceptible de conduire à l'efficacité productive soit le duopole : ainsi il serait « souhaitable » que la production de l'ensemble des unités de l'output soit répartie entre deux entités de production. On parlerait alors de duopole naturel (et de la même manière, il pourrait exister des situations de triopole naturel, de quatropole naturel, etc.). Établir une telle typologie des configurations « naturelles » de marché n'est en réalité ni commode, ni suffisant pour établir avec certitude quelle configuration apporterait le plus fort niveau de bien-être social à la collectivité (ceci dépasse le cadre de cet ouvrage). Nous nous contenterons de retenir que :

- les conditions technologiques ne sont pas neutres sur le type de structure de marché qui rend l'obtention de l'efficacité sociale envisageable ;
- en dépit du caractère non coopératif du jeu, le duopole de Cournot ne conduit pas à une situation aussi favorable pour les consommateurs que le cadre de la concurrence pure et parfaite ;
- la recherche d'une solution coopérative (si elle est légale et/ou possible) devrait permettre aux firmes d'obtenir conjointement un profit plus élevé.

Avant de discuter de ce que serait la recherche, par les firmes, d'une solution coopérative, nous allons envisager le cas, dans une optique non coopérative, d'un duopole non symétrique.

3 Le duopole de Stackelberg

Dans la section précédente, consacrée à un duopole symétrique « en quantité », nous avons suggéré que la présence de deux firmes « telles qu'aucune ne possède un ascendant sur l'autre » ne pouvait se concevoir qu'en envisageant des structures de tailles, de technologies et d'anciennetés relativement comparables. Ceci n'est pas toujours ce que l'on observe sur les marchés de biens et services. Par exemple, telle boisson pétillante produite et distribuée de fort longue date à travers le monde est en position de leader relativement à ses concurrentes, plus petites, plus récentes ou de plus faible notoriété. Tel fabricant de chariots élévateurs (pour l'industrie) est dans une position tellement dominante sur le marché que l'on confond parfois la désignation du produit et la marque de ce leader. Tel fabricant de polos (pouvant être portés sur un court de tennis) a une notoriété telle qu'il conditionne la stratégie mimétique de la masse de ses suiveurs. Les exemples sont nombreux (lunettes de soleil, bottes de cuir façon cow-boy, motocyclettes pétaradantes à long guidon, etc.). Les firmes en position de leader ont la capacité à obtenir un profit de long terme supérieur à celui des suiveurs (les « followers »). C'est cette intuition que se propose d'étudier le duopole de Stackelberg.

Le **duopole de Stackelberg** est une configuration de marché dans laquelle deux firmes en positions asymétriques se font concurrence en quantité. Elles produisent et vendent un bien homogène à un prix unique p. L'une des firmes est dominante : c'est le leader. L'autre est dominée : c'est le follower. Chacune connaît les caractéristiques de la relation entre prix et quantité demandée, les caractéristiques de sa propre fonction de coût ainsi que celles de la fonction de coût de sa concurrente.

Une fois encore, le cadre dans lequel nous envisageons d'analyser la formation des décisions des firmes leader et follower est un cadre qui semble assez éloigné de la réalité : la toute première variable stratégique serait plus naturellement le prix et l'idée même d'un bien homogène paraît encore moins raisonnable dans le cadre de Stackelberg que dans le cadre de Cournot. Néanmoins, la caractérisation de l'équilibre non coopératif obtenu est valable, qualitativement très proche de ce qui se passerait dans les circonstances réelles.

Le jeu de Stackelberg est un jeu séquentiel à 2 étapes. Lors d'une première étape, la firme leader (nous postulons qu'il s'agit de la firme 1) décide de sa production Q_1^S en maximisant son profit. Dans une seconde étape, la firme follower s'adapte à la décision du leader en se positionnant sur sa fonction de meilleure réponse : $Q_2^S = MR_2\ (Q_1^S)$. Pour déterminer les caractéristiques de cet équilibre non coopératif $(Q_1^S\ ;\ Q_2^S)$, il faut résoudre le jeu par la méthode dite de la « récurrence à rebours ». Ce terme, laide traduction de l'expression anglaise « backward induction », désigne le mode de résolution de tous les jeux séquentiels : il faut partir de la fin du jeu puis remonter, pas à pas, vers le début du jeu. Dans le cas présent, il faut donc commencer par l'étude de la décision de la firme 2 (le follower), puis intégrer les caractéristiques de la réponse optimale de la firme 2 dans l'ensemble d'informations de la firme 1, le leader ; sur la base de ces informations, la firme leader construit alors sa décision optimale (de maximisation du profit). Le résultat obtenu est différent de celui obtenu dans le cadre symétrique de Cournot : la firme leader dispose d'un avantage incarné par l'opportunité de jouer en premier. La firme follower joue en second et adopte un comportement optimisateur, mais elle s'adapte « passivement » à la décision du leader : elle ne conteste pas son leadership. Bien sûr, le jeu ici décrit n'est qu'un habillage théorique de la réalité ; il se pourrait en effet, que, dans la vie réelle, les deux firmes agissent simultanément (ou même que le follower prenne sa décision un peu avant le leader). Mais la vertu de cette modélisation est d'objectiver, dans le cadre d'un jeu parfaitement délimité, l'émergence d'un comportement de leader et d'un comportement de *follower* sur un marché. L'équilibre non coopératif du duopole de Stackelberg est représenté sur la figure 5.

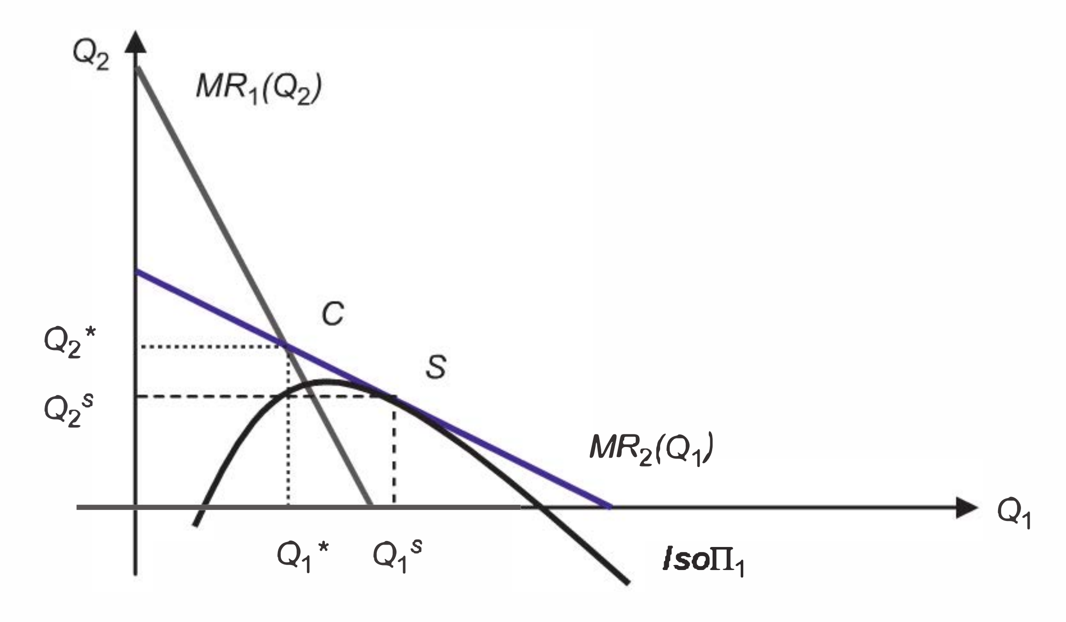


Figure 5 Équilibre non coopératif du duopole de Stackelberg

L'équilibre non coopératif de Stackelberg est le couple stratégique $(Q_1^S ; Q_2^S)$ tel qu'en ce point la tangente à la courbe d'iso-profit de la firme 1 soit confondue avec la fonction de meilleure réponse de la firme 2. C'est en effet la solution graphique qui correspond à l'issue du jeu séquentiel décrit ci-dessus.

À fonctions de coûts identiques, la firme leader produit, à l'équilibre non coopératif de Stackelberg, plus que la firme follower. Ainsi, le leader réalise un profit plus élevé que le follower. Si l'on compare, en outre, les duopoles de Cournot et de Stackelberg à fonctions de coût identiques, on constate que la quantité produite par la firme leader est plus élevée que la production de l'une ou l'autre des concurrentes dans le cadre de Cournot. *A contrario*, le follower produit moins. On peut aussi constater sur la figure 6 que l'équilibre non coopératif de Stackelberg n'est, pas plus que l'équilibre de Cournot, un optimum de Pareto : une fois encore, la présence d'une lentille contenant des couples de productions qui permettraient aux firmes d'accroître conjointement leurs profits matérialise le caractère non Pareto-optimal de l'équilibre non coopératif du jeu de Stackelberg.

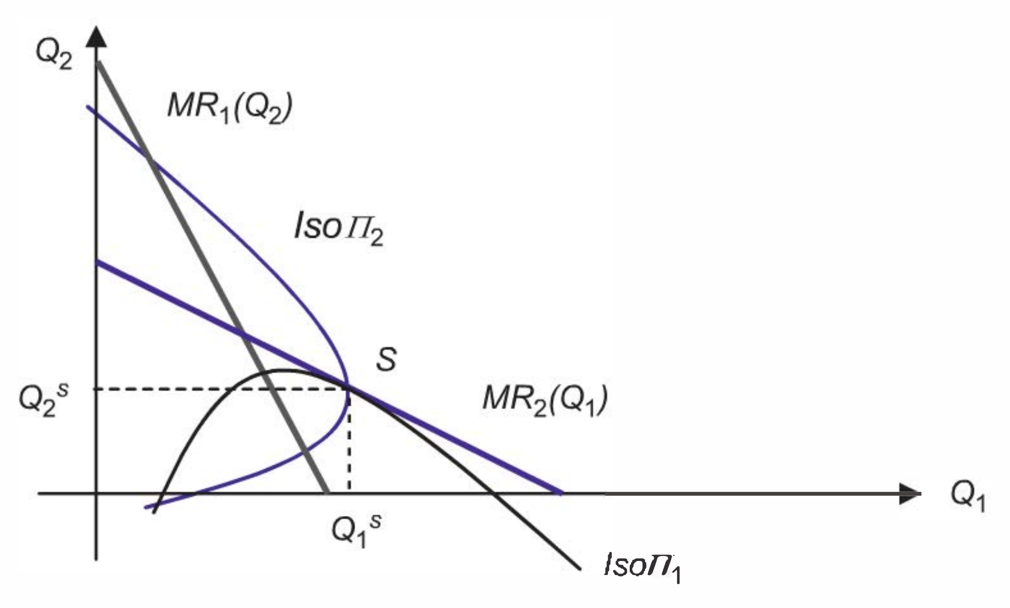


Figure 6 L'équilibre non coopératif du duopole de Stackelberg n'est pas un optimum de Pareto

Pour aller plus loin

Détermination de l'équilibre de Stackelberg

Nous connaissons déjà la fonction de meilleure réponse de la firme 2, $Q_2 = MR_2\ (Q_1)$, issue de la maximisation du profit de la firme 2. C'est, en tout cas, ce qui doit être déterminé lors de la première étape de résolution du jeu.

Lors de la seconde étape de la résolution, il s'agit de caractériser la décision de la firme 1 qui maximise son profit sachant que la firme 2 se positionnera sur sa fonction de meilleure réponse. Ceci revient à :

$$\underset{Q_1}{Max}\ \Pi_1\ (Q_1\ ;\ Q_2) \text{ sous contrainte : } Q_2 = MR_2\ (Q_1)$$

Pour résoudre ce programme de maximisation, il faut substituer, dans la fonction objectif, Q_2 par l'expression de la meilleure réponse de la firme 2 à la stratégie de la firme 1 :

$$\underset{Q_1}{Max}\ \Pi_1\ (Q_1) = p[Q_1 + MR_2\ (Q_1)] \times Q_1 - C_1(Q_1)$$

La condition d'optimalité est que $\frac{\partial \Pi_1(Q_1)}{\partial Q_1} = 0$. La valeur de Q_1 qui annule cette dérivée première est $Q_1{}^S$. Enfin, $Q_2{}^S = MR_2\ (Q_1{}^S)$.

4 Discussion sur la variable stratégique : le duopole de Bertrand

Le choix de la quantité plutôt que du prix comme variable stratégique a suscité de nombreuses critiques. Celles de Joseph Bertrand furent les plus virulentes. À juste titre, il convient de s'interroger sur ce qui se passerait si les firmes, plutôt que décider des quantités à produire, concentraient leur réflexion stratégique sur le prix à pratiquer. Cette question est un véritable défi pour la construction de Cournot : si le prix était la variable stratégique, les firmes finiraient, à l'issue d'une « guerre des prix » par tarifer le bien ou service produit au coût marginal, comme en concurrence pure et parfaite !

Présentons, succinctement, le raisonnement de Bertrand. Supposons que les deux firmes ont des fonctions de coût semblables, et donc les mêmes fonctions de coût marginal ; supposons de surcroît que le coût fixe est nul. Quel que soit le prix p pratiqué par l'une des firmes, sa concurrente aura intérêt à pratiquer un prix $p - \varepsilon$ (où ε désigne une valeur infime positive) lui permettant d'attirer à elle tous les clients (et d'accroître son profit). La première firme ne resterait pas sans réaction et pratiquerait alors le prix $p - 2\varepsilon$ lui garantissant à son tour d'attirer tous les acheteurs. Mais la seconde firme réagirait et pratiquerait le prix $p - 3\varepsilon$... et ainsi de suite. Cette guerre des prix s'arrêterait à un plancher correspondant au coût marginal de production de la moitié de la quantité globalement produite. Ainsi, les firmes parviendraient à un équilibre non coopératif dans lequel chacune produirait la moitié des unités d'output et vendrait ces unités d'output au prix plancher égal à leur coût marginal de production. Elles réaliseraient toutes deux un profit microéoconomique nul et la situation du marché s'apparenterait à une situation conforme aux conditions de la concurrence pure et parfaite.

Il y a deux manières de réagir au raisonnement de Bertrand. Soit jeter le travail de Cournot aux oubliettes (ce qui, heureusement, n'arrivera pas), soit comprendre que l'hypothèse d'un bien homogène ne permet pas (c'est une évidence) de raisonner en supposant que la variable stratégique est le

prix. Il faudrait modéliser la différenciation des produits pour être à même de travailler avec le prix comme variable stratégique. C'est l'objet des modèles de concurrence monopolistique, et en particulier des modèles de localisation spatiale comme les modèles de Hotelling et de Salop ; la modélisation des « distances » est en effet une démarche impérative pour analyser comment des consommateurs font leur choix entre des produits différenciés : « distances » entre les caractéristiques des différents produits et « distances » entre les caractéristiques des produits et les goûts des consommateurs. Dans ces modèles, à prix identiques des biens différenciés, les consommateurs choisissent les produits qui leur sont les plus proches, au sens propre (géographiquement) ou figuré (distance entre caractéristiques).

Mais il y a aussi un enseignement à tirer de l'analyse de Bertrand : la rivalité concurrentielle entre les firmes peut être exacerbée et conduire à une issue peu différente de ce qui se produirait en situation de concurrence pure et parfaite. On parle d'ailleurs du **paradoxe de Bertrand** : alors que les firmes sont peu nombreuses et de tailles conséquentes, la concurrence entre celles-ci les conduit à se comporter comme si elles étaient très nombreuses et de petites tailles. On peut en conclure qu'une loi qui lierait le prix d'équilibre au nombre de firmes présentes sur le marché (de la forme : le prix d'équilibre du marché est d'autant plus élevé que le nombre de firmes présentes sur le marché est faible) ne serait pas très pertinente. Ce que nous pouvons conclure est que d'autres paramètres, non pris en compte dans la modélisation rudimentaire proposée, sont susceptibles de conditionner le caractère plus ou moins exacerbé de la rivalité concurrentielle.

5 La collusion

À l'occasion de l'étude des duopoles de Cournot et Stackelberg, nous avons montré que les équilibres non coopératifs atteints n'étaient pas des optima de Pareto. Dans chacun des cas, la présence d'une lentille contenant les couples de productions (Q_1 ; Q_2) qui accroîtraient conjointement le profit des protagonistes matérialisait l'existence d'opportunités d'ententes entre les firmes. À l'opposé de l'optique non coopérative, il peut en effet exister des situations où les firmes s'entendent au préalable : c'est la collusion.

La **collusion** désigne une situation où, sur un marché, deux ou plusieurs firmes s'entendent pour fixer le prix de vente et/ou la quantité produite de leurs outputs. Une entente peut regrouper toutes les firmes présentes sur le marché ou un sous-ensemble seulement. En microéconomie, les termes collusion, coopération, cartel ou entente sont synonymes.

La recherche, par deux ou plusieurs firmes, d'un équilibre coopératif n'est, en général, pas légale. Les autorités interdisent les ententes dans la mesure où celles-ci dégradent le *welfare* : la perte de surplus enregistrée par les consommateurs est supérieure au supplément de profit que peuvent obtenir les firmes. De manière un peu paradoxale, le caractère illégal des ententes peut contribuer à ce qu'une criminalité émerge pour les faire respecter, car il n'est pas possible d'aller devant les tribunaux pour faire valoir son « bon droit ». Si une entente secrète est conclue, en faire respecter les termes est très difficile ; l'un des seuls moyens d'y parvenir est l'intimidation, la menace… ou pire encore. Nous allons voir ci-dessous que chaque participant à une entente aurait intérêt à dévier unilatéralement des termes de l'entente et que l'interrogation sur les moyens à mettre en œuvre pour dissuader de tels comportements n'est pas une question anecdotique.

Essayons tout d'abord d'établir la position des « contrats » (illégaux) que pourraient passer les deux firmes désireuses de s'entendre pour accroître conjointement leurs profits. Appuyons-nous sur les courbes d'iso-profit de ces deux firmes. Nous pouvons identifier des contrats particuliers qui ont la caractéristique d'être, pour les firmes, des conventions de partage de la production telles que toute modification des termes du partage au bénéfice de l'une ne puisse se faire qu'au détriment de l'autre. Nous reconnaissons là des couples de quantités produites qui ne peuvent être que des **points de tangence** entre les courbes d'iso-profit des deux firmes (comme, dans le chapitre 5, les points de tangence entre les courbes d'indifférence des deux agents étaient les allocations telles qu'aucun individu ne pouvait accroître sa satisfaction sans détériorer celle de l'autre). Faisons apparaître, sur la figure 7, ces contrats tels qu'aucune firme ne puisse accroître son profit sans détériorer celui de sa concurrente.

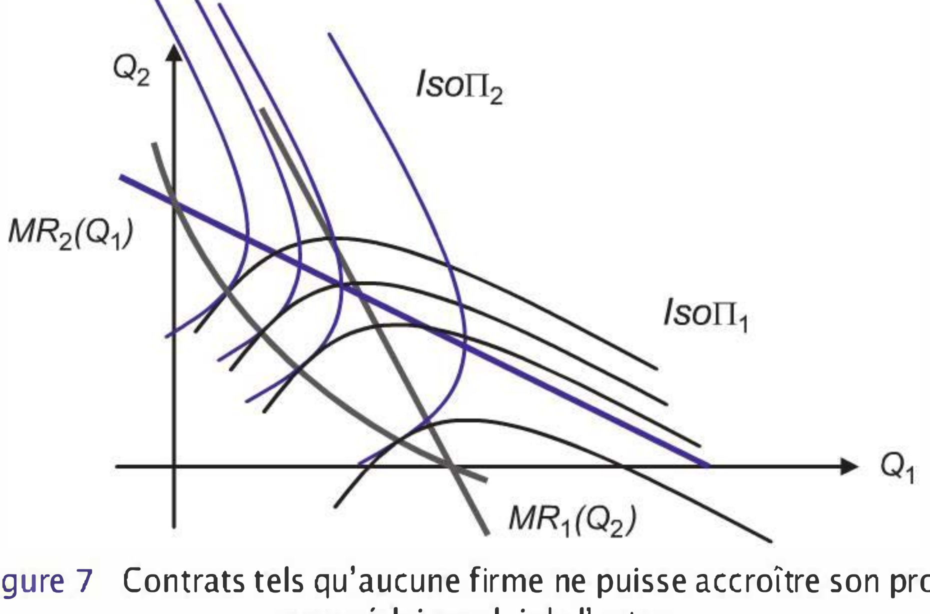


Figure 7 Contrats tels qu'aucune firme ne puisse accroître son profit sans réduire celui de l'autre

Nous voyons ainsi apparaître une courbe (en gras) qui caractérise les contrats qui sont envisageables à l'occasion d'une entente. En réalité, seule une petite partie de cette courbe décrit le domaine sur lequel les couples de quantités produites sont susceptibles de faire l'objet d'un accord : seuls les contrats apportant un profit supérieur à celui obtenu dans le cadre de l'équilibre non coopératif (symétrique) sont dignes d'intérêt pour les deux protagonistes. Il faut donc restreindre cet arc à la zone contenue entre les courbes d'iso-profit atteintes lorsque les firmes sont à l'équilibre non coopératif du duopole (de Cournot). Nous représentons cette zone appelée « cœur » de l'entente sur la figure 8. Tous les points formant le petit arc en gras sont des contrats susceptibles de constituer les termes de l'entente. Reste à établir quel contrat sera choisi. *A priori*, seul le rapport de force entre les firmes détermine l'issue de la négociation. On dit alors que c'est la confrontation entre les **pouvoirs de négociation** de chacune des firmes qui décide des termes exacts de l'entente (c'est-à-dire, ici, la quantité totale produite et la répartition de cette quantité produite entre les protagonistes). Dans certains cas, les contraintes technologiques (la forme des courbes de coût) peuvent conduire les firmes à convenir d'une répartition quasi « mécanique » de la production : le « rapport de force » ne joue alors presqu'aucun (ou plus aucun) rôle dans l'établissement des termes de l'entente.

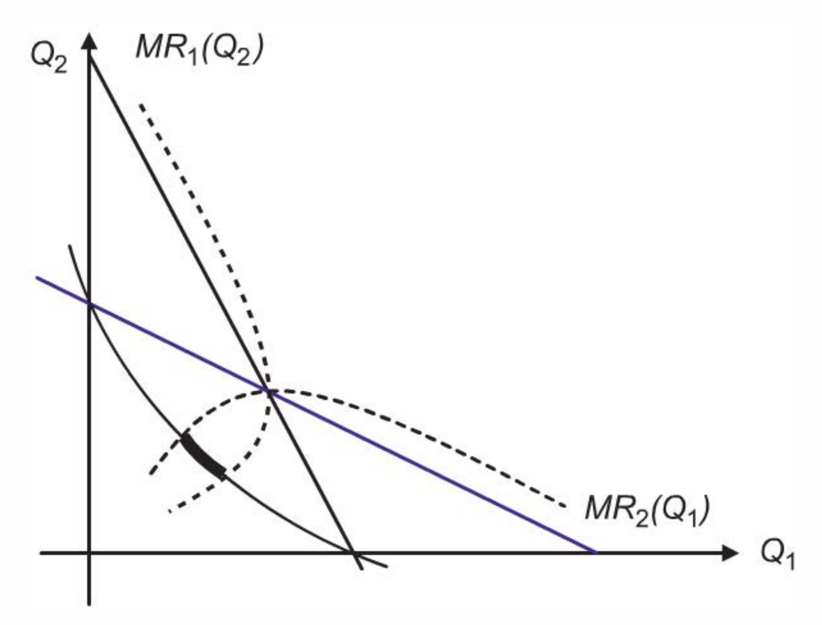


Figure 8 « Cœur » de l'entente

Représentons, sur la figure 9, un contrat d'entente (Q_1^E ; Q_2^E) choisi, à l'issue d'une négociation, dans le cœur.

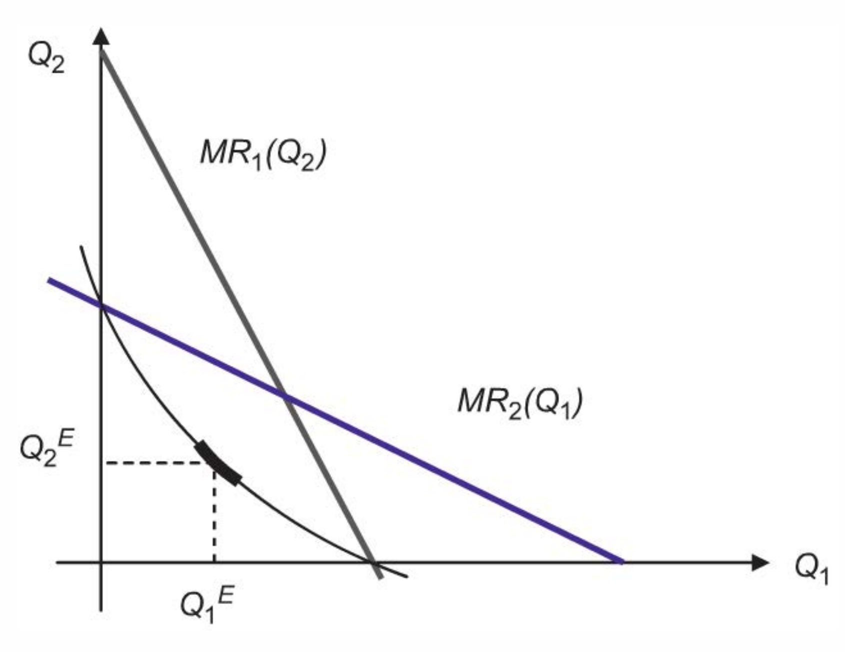


Figure 9 Termes d'une entente

Il est opportun de comprendre qu'un tel contrat est intrinsèquement instable au sens où, chaque firme a, à tout instant, intérêt à dévier unilatéralement des termes de son engagement. Certes, le contrat (Q_1^E ; Q_2^E) qui accroît le profit des deux firmes (relativement à la situation d'équilibre non coopératif) a fait l'objet d'un accord entre elles, mais chacune sait

(ou établit) qu'elle pourrait encore accroître son profit en se positionnant sur sa fonction de meilleure réponse relativement à la quantité que s'est engagée à produire sa concurrente. C'est ce que nous illustrons sur la figure 10.

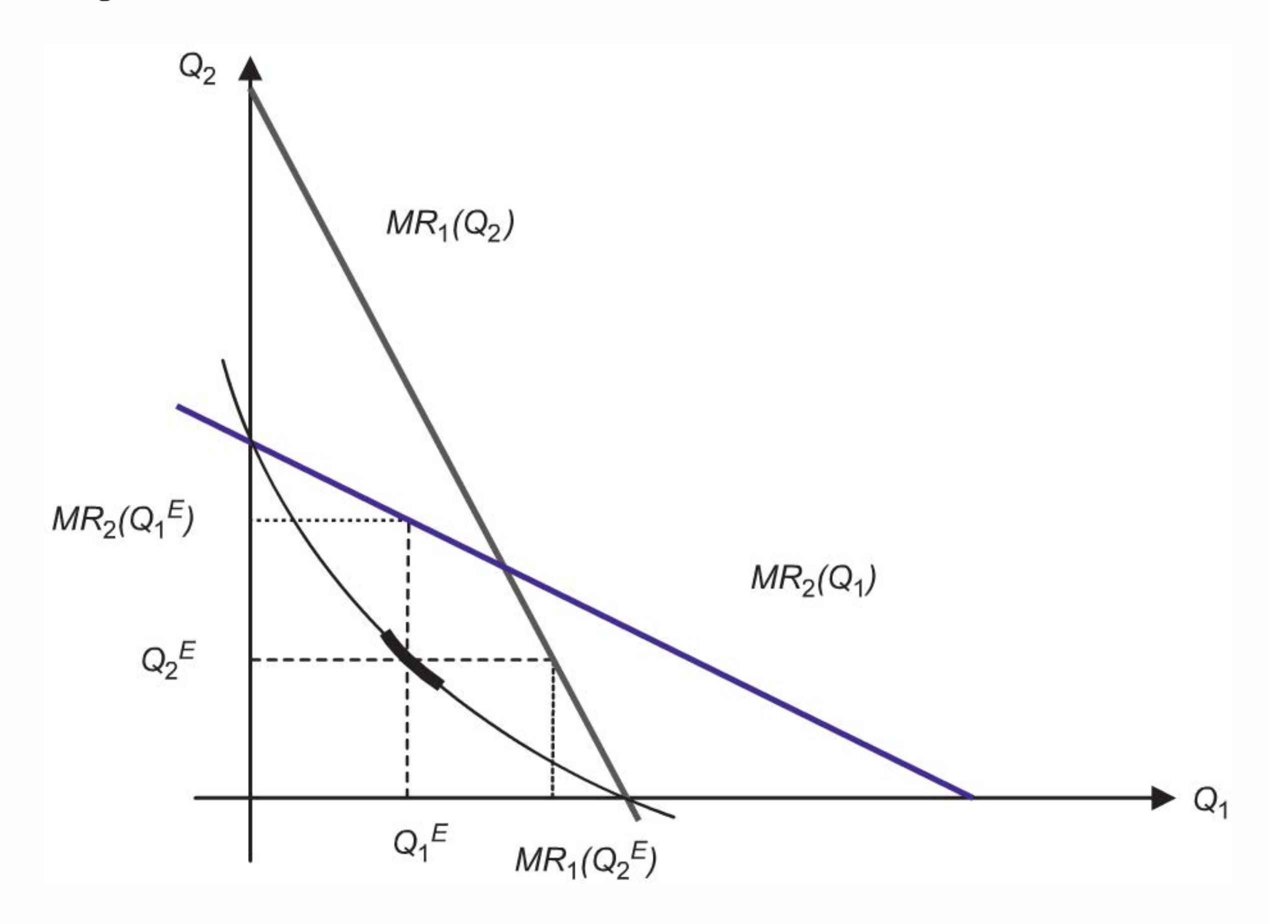


Figure 10 Chaque firme a intérêt à dévier unilatéralement de son engagement

On le comprend, toute entente est donc fragile. Dans une optique intertemporelle, néanmoins, il peut être opportun, pour chaque firme, de respecter les termes de son engagement car toute déviation entraîne, à la période suivante, des « représailles » économiques de la part de son adversaire : on peut alors avoir plus à perdre sur les périodes futures que ce que l'on escompte gagner lors de la période présente.

Pour aller plus loin

Détermination des termes de l'entente

Le traitement analytique de la collusion consiste à supposer que les firmes maximisent un profit joint. Pour identifier les termes $(Q_1^E ; Q_2^E)$ de l'entente, il s'agit donc de :

$$\underset{Q_1, Q_2}{Max} \ \Pi(Q_1 + Q_2) = p(Q) \times Q_1 + p(Q) \times Q_2 - C_1(Q_1) - C_2(Q_2)$$

Les conditions d'optimalité sont que : $\begin{cases} \dfrac{\partial \Pi(Q_1 + Q_2)}{\partial Q_1} = 0 \\ \dfrac{\partial \Pi(Q_1 + Q_2)}{\partial Q_2} = 0 \end{cases}$

Selon les caractéristiques des fonctions de coût, la solution de ce système d'équations sera soit un unique contrat $(Q_1^E ; Q_2^E)$, soit un ensemble de contrats telle que la quantité $Q_1^E + Q_2^E$ soit parfaitement identifiée. Dans ce second cas, c'est le pouvoir de négociation de chaque firme qui déterminera les termes exacts de l'entente.

11 Notions d'économie publique

Mots-clés

Effets externes, taxe pigouvienne, droits à polluer, biens publics, passager clandestin

Dans le chapitre 5, nous avons indiqué que le résultat central en vertu duquel l'équilibre général d'une économie concurrentielle est socialement efficace, est relativisé en présence de biens collectifs et/ou d'externalités. En d'autres termes, dans ces cas, le libre jeu du marché peut ne pas conduire à l'efficacité paretienne (au sens de Pareto) et l'intervention de la puissance publique (sous une quelconque forme) s'avère nécessaire. Ce résultat de la théorie microéconomique légitime *a posteriori* des pratiques bien ancrées : la défense de l'entité territoriale et la sécurité publique sont assurées par des organisations placées sous l'autorité du souverain depuis la naissance des civilisations. De même, des systèmes d'évacuation des eaux usées ont été mis en place par les cités depuis l'époque romaine ; l'éclairage des rues a été systématisé dans les villes depuis le XVII[e] siècle… En présence de biens ou services collectifs, les règles qui président à la production et à l'échange des biens diffèrent de celles que l'on rencontre pour les biens privés ; l'étude de ces règles est l'objet de ce chapitre. Mais la frontière entre bien privé et bien collectif est parfois mince et il n'est pas toujours simple d'établir si la fourniture du bien doit être assurée par des financements privés ou des fonds publics. Il existe, d'autre part, des biens et services dont la production affecte la satisfaction d'agents autres que les producteurs et les consommateurs du bien en question. Ce type de phénomènes, appelé effet externe (ou externalité) – et dont la pollution est un exemple emblématique – peut lui aussi nécessaire rendre nécessaire une intervention de la puissance publique.

1 L'existence d'effets externes et leur internalisation

Lorsqu'un agent économique quelconque prend une décision, il engendre potentiellement des changements du niveau auquel se fixent les prix dans toute l'économie : on peut imaginer que les prix de tous les biens et services soient affectés, quand bien même les variations de la plupart des prix sont infinitésimales. Ainsi, d'un point de vue théorique, le sort de tous les agents de l'économie est susceptible d'être affecté par l'action d'un seul. Tant que la perturbation de l'économie est pleinement internalisée par le système des prix, il n'y a pas d'effets externes. En revanche, dès qu'une action entraîne des conséquences dont le signal prix ne parvient pas à informer de manière adéquate les différents acteurs de l'économie, on parle d'effet externe.

On appelle **effet externe** ou **externalité** tout phénomène dont l'origine est une action d'un agent économique affectant la satisfaction ou l'ensemble de choix d'un ou plusieurs autres sans que les éventuelles variations de prix conséquentes à cette action ne véhiculent convenablement l'information sur l'état de l'allocation des ressources dans l'économie.

Pourquoi une décision peut-elle ne pas être convenablement internalisée par le système de prix ? En raison de la différence entre le coût privé et le coût social (ou entre le bénéfice privé et le bénéfice social) d'une activité productive. Certaines actions économiques ont, en effet, un coût social supérieur au coût privé. Prenons l'exemple d'une entreprise chimique qui déverserait ses déchets dans le cours d'une rivière. Pour l'entreprise, le coût privé de traitement des déchets serait alors nul. Pour les habitants en aval, alimentés en eau potable à partir d'une captation dans la rivière, le coût social est, en revanche, élevé : si l'ingestion de l'eau provoque des allergies, maladies, empoisonnements, etc., il faudra se soigner (ce qui est coûteux au sens pécuniaire du terme) et/ou perdre quelques années de vie en bonne santé (ce qui est coûteux au sens d'une perte d'opportunités). Le coût

social est le coût d'une action économique pour la collectivité dans son ensemble. Dans l'exemple précédent, on comparera le coût de production de la substance fabriquée par l'entreprise chimique (et supportée par elle seule), que l'on appelle donc le coût privé, et le coût social de la production de cette substance, qui inclut, outre le coût de fabrication à proprement parler, le coût des nuisances pour les riverains victimes d'une dégradation de la qualité de l'eau. Puisqu'ici, le coût privé est inférieur au coût social, on parlera d'externalité négative. Quand le bénéfice social est supérieur au bénéfice privé, on parle d'externalité positive.

La question que se pose l'économiste n'est pas tant de savoir si les pratiques de l'entreprise chimique doivent cesser ou sont moralement condamnables, mais de trouver un moyen pour faire coïncider le coût privé et le coût social. Ainsi, il se peut que le dispositif qu'il suggérera de mettre en place, et qu'il considérera comme économiquement satisfaisant, ne parvienne qu'à réduire partiellement – et non totalement – les rejets toxiques dans la rivière. L'économiste ne se place pas sur le terrain de la morale, il ne se préoccupe que d'efficacité économique (au sens de la Pareto-optimalité). C'est pourquoi l'économiste parle volontiers de « niveau de pollution optimal », sans que ce niveau ne soit nécessairement nul. Ainsi pourra-t-il trouver optimal que l'entreprise chimique, soumise à la taxe que nous allons présenter ci-dessous, rejette moitié moins de déchets toxiques dans l'eau et que le produit de la taxe soit utilisé pour indemniser les riverains (qui l'utiliseront pour soigner leurs allergies ou maladies)…

Le premier mode d'internalisation des externalités est la mise en place d'une taxe. Cette taxe est dite pigouvienne, en référence au grand économiste Arthur C. Pigou. On parle aussi parfois, de manière plus équivoque, de taxe « pollueur-payeur » ou d'éco-taxe. Une telle taxe doit être fixée à un niveau permettant de faire coïncider le coût supporté par l'entreprise incriminée (coût privé + taxe) et le coût social. Plus précisément, ce sont les coûts marginaux, privé et social, qui sont en jeu. Sur la figure 1, sont représentées les courbes de coût marginal privé et de coût marginal social relatifs à une entreprise dont l'activité est à l'origine d'une externalité négative. Le montant de la taxe pigouvienne apparaît comme la différence entre les deux coûts marginaux.

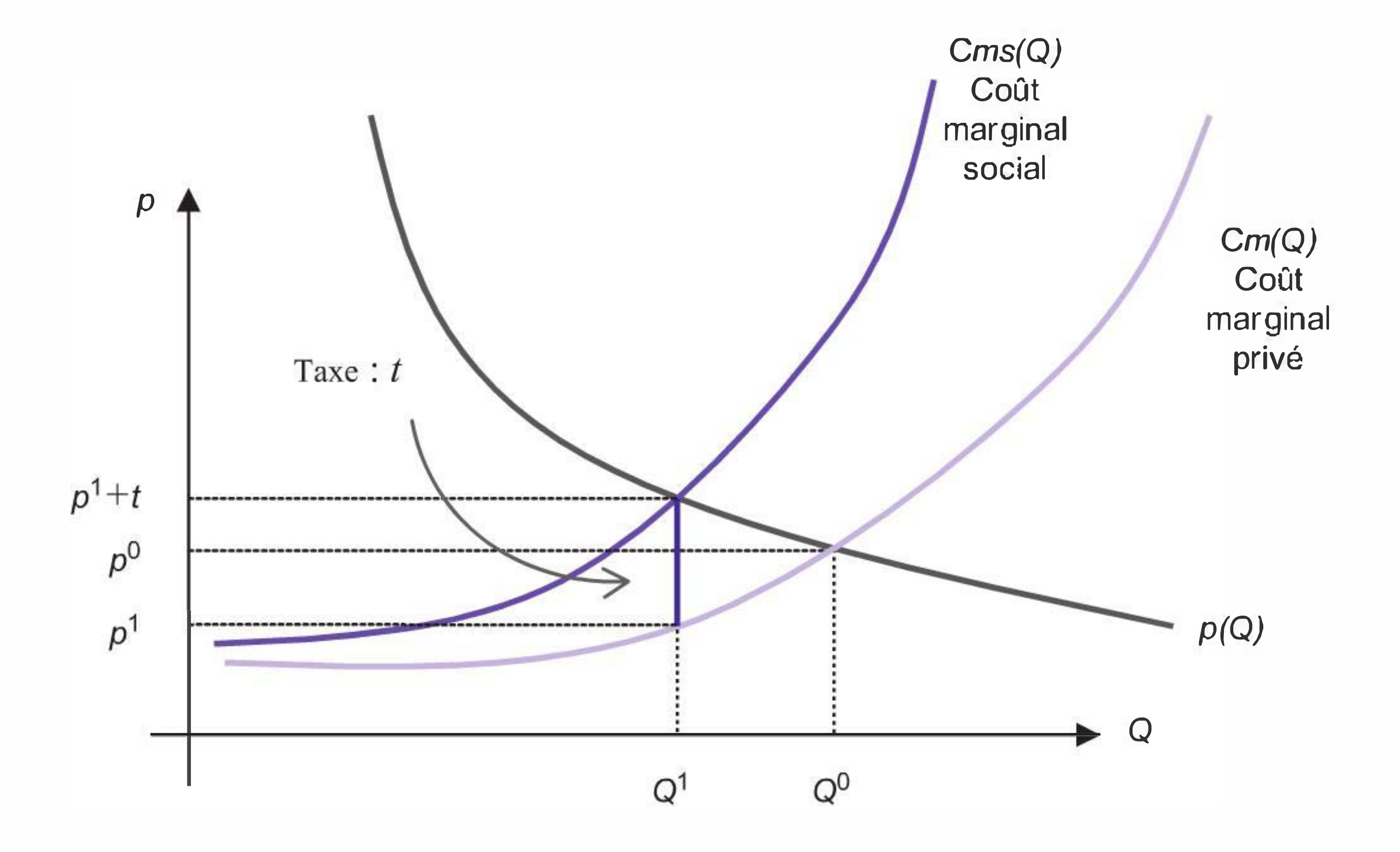


Figure 1 Taxe pigouvienne

Le couple $(Q^0 ; p^0)$ est le couple quantité-prix, à l'intersection de la demande inverse et du coût marginal, qui correspondrait à l'optimum social en l'absence d'externalité (en l'absence d'écart entre le coût marginal social et le coût marginal privé). La présence de l'externalité fait que ce couple $(Q^0 ; p^0)$ ne permet plus d'atteindre l'optimum social. C'est pourquoi la tutelle impose une écotaxe de niveau t (par unité produite) qui pousse la firme à réduire sa production au niveau $Q^1 < Q^0$. Sous les hypothèses retenues (coûts marginaux croissants), le prix unitaire demandé au consommateur sera en baisse : $p^1 < p^0$. Le produit de la taxe $t \times Q^1$ sera perçu par la puissance publique et redistribué aux agents dont la satisfaction a été dégradée par l'activité polluante. Si nous reprenons l'exemple de l'entreprise chimique qui pollue une rivière, nous devinons que la pollution sera réduite mais ne sera sans doute pas totalement jugulée. Ce qui importe est que l'indemnisation monétaire des riverains compense la désutilité engendrée par la pollution résiduelle.

On appelle **taxe pigouvienne** une taxe égale à l'écart entre le coût marginal privé et le coût marginal social de l'activité productive. Cette taxe permet, en théorie, de restaurer l'efficacité sociale.

Une manière moins extrême d'envisager cette solution pigouvienne est de supposer que, face à la menace de la taxe, l'entreprise va investir dans des activités de dépollution ou de traitement des déchets. Dès lors, son coût marginal privé augmente (puisqu'il est désormais égal à la somme du coût marginal de production et du coût marginal de dépollution) et se rapproche du coût marginal social. Ainsi la taxe diminue et peut même devenir nulle si les deux coûts marginaux coïncident.

Une telle taxe serait un dispositif très satisfaisant en théorie mais compliqué à mettre en œuvre en pratique. D'abord parce qu'il est difficile, pour la puissance publique, de connaître la fonction de coût (de l'entreprise ou) des entreprises incriminées. Ensuite, parce qu'il est difficile de déterminer qui sont les victimes de l'externalité et à combien s'élève le niveau de la désutilité qu'elles subissent.

Un autre dispositif permet, en théorie, de corriger les effets indésirables des externalités : ce sont les permis d'émission négociables, plus couramment dénommés « droits à polluer ». Pour comprendre le dispositif des droits à polluer, il faut se placer dans une optique un peu différente de celle de Pigou. L'idée centrale, développée par Ronald Coase, est que le dysfonctionnement auquel on assiste est dû à une mauvaise allocation des droits de propriété sur les différentes « ressources ». Dans la logique coasienne, l'inefficacité sociale engendrée par la présence d'externalités peut disparaître si les différents acteurs peuvent acheter et vendre les droits à polluer (l'air, l'eau, etc.) sur un marché : spontanément, et quelle que soit la répartition initiale des droits de propriété entre les pollueurs et les pollués, le libre jeu du marché (où les acheteurs achètent librement et les vendeurs vendent librement) devrait conduire à une allocation des ressources pour laquelle sera restaurée l'efficacité sociale. Ce résultat est connu comme le « théorème de Coase ».

Théorème de Coase

En présence d'externalités, un système de libre marchandage des droits de propriétés sur les ressources conduit à l'efficacité économique, quelle que soit la répartition initiale des droits de propriété entre les différents acteurs du marché.

Notons, bien sûr, que l'optimum de Pareto atteint différera selon la position des dotations initiales. Il ne s'agit là que de juger du cheminement vers l'efficacité parétienne (et non vers une allocation « juste » ou équitable des ressources). Si les droits à polluer sont initialement alloués aux consommateurs (aux « pollués » pour schématiser), les firmes qui désirent émettre des tonnes de CO_2 devront leur acheter ces droits : si, en début de période, les firmes sont désireuses d'émettre globalement plus que le total des émissions disponibles (demande de droits à polluer supérieure à l'offre), les prix des permis augmentent et seules les firmes dont les disponibilités maximales à payer sont les plus élevées (car proportionnelles aux perspectives de profitabilité associées à l'activité productive) parviennent à se doter des droits. Une solution alternative pour les firmes est d'investir dans des technologies moins polluantes ou dépolluantes, puisque, au regard du prix élevé des permis, ces solutions deviennent économiquement opportunes. En outre, si les consommateurs, initialement détenteurs des droits à polluer, désirent ne céder qu'une partie des droits dans l'optique de privilégier la qualité de l'air et des ressources environnementales, le prix des permis continuera à augmenter (seules les entreprises les plus profitables ou celles dont l'activité est la plus crucialement dépendante de l'autorisation d'émettre les substances polluantes enchériront suffisamment pour acquérir le petit stock de droits mis en vente). Pour les consommateurs détenteurs des droits, il faut alors un très grand attachement à la qualité des ressources environnementales pour résister aux opportunités monétaires qu'offrent la possible vente des permis.

Cette remarque nous amène à une question centrale qui est de savoir à qui sont attribués les permis, par quelle tutelle et pour quel niveau d'émission ? Définir un quota annuel d'émission est très complexe : les experts s'entendent difficilement sur ce qu'est le sentier de croissance soutenable

(c'est-à-dire préservant la qualité des ressources environnementales). En outre, la pollution se déplace et se répand sans se préoccuper des frontières ce qui rend nécessaire une coordination internationale s'agissant de définir un quota d'émission continental ou planétaire. Quant à l'identité, par décision de la tutelle, des attributaires initiaux des permis, il s'agit d'un choix politique aux implications de « justice sociale » conséquentes. Par ailleurs, l'expérience montre que les marchés existants des droits à polluer ne fonctionnent pas toujours très bien : soit les marchés sont inactifs pendant de longues périodes, soit ils sont en proie à des mouvements spéculatifs (des investisseurs achètent les permis en début de période et les conservent en portefeuille ; les entreprises souhaitent acquérir des permis pour produire tout au long de l'année, mais les investisseurs « retiennent » les permis pour en faire grimper les prix ; une fois que les prix des permis se sont envolés, les investisseurs les revendent et empochent une plus-value). Enfin, la présence de coûts de transaction de toute nature est susceptible d'affaiblir la validité du théorème de Coase.

2 L'analyse économique des biens publics

Les biens dont nous avons étudié la production, la consommation et l'échange dans les chapitres précédents étaient implicitement des biens privés. Les biens privés sont tous les biens qui vérifient simultanément deux propriétés : la rivalité et l'exclusivité. Par rivalité, il faut entendre que, dès lors qu'un agent consomme une unité d'un bien (un fruit, par exemple), cette unité de bien n'est plus consommable par un autre agent. Par exclusivité, il faut entendre que, dès lors qu'un agent a acquitté le prix d'achat du bien, ce bien n'est plus disponible et/ou accessible pour un autre agent.

> Les **biens privés** sont des biens et services qui vérifient les propriétés de rivalité et d'exclusivité.

Dès lors qu'une des deux propriétés, rivalité ou exclusivité, n'est pas vérifiée, il s'agit de biens publics ou collectifs. Plusieurs typologies existent.

Certains biens publics réunissent des spécificités qui les rangent à l'exact opposé des biens privés purs. Les trois spécificités qui caractérisent les biens publics purs sont : l'impossibilité d'exclusion par l'usage, l'obligation d'usage et l'absence d'effets d'encombrements.

- L'impossibilité d'exclusion par l'usage signifie que l'on ne peut exclure un tiers de la jouissance de l'usage du bien en raison de sa propre consommation. L'exemple type est celui de l'éclairage public : ce n'est pas parce que je jouis de ce que les rues de la ville sont éclairées la nuit que je prive un autre piéton de ce confort. Un autre exemple est celui de la défense nationale. Ce n'est pas parce que l'arsenal nucléaire de mon pays (qui dissuade une puissance belliqueuse de venir, jusque dans nos campagnes, égorger nos fils et nos compagnes) me protège qu'il ne protège pas (ou protège moins) mon voisin.
- Par l'obligation d'usage, chaque agent est tenu de consommer le bien ou de recourir au service. L'école obligatoire (jusqu'à 16 ans) est un service public assorti d'une obligation d'usage : la société estime en effet que la scolarisation de tous les enfants est nécessaire à la connaissance du contrat social et à une adhésion minimale à ses termes. La justice est aussi un service assorti d'une obligation d'usage (en cas de survenance d'un crime ou délit).
- L'absence d'effets d'encombrements existe d'autant plus que les biens sont immatériels. Par exemple, l'opportunité de jouir, par le biais d'un *tuner* traditionnel, des programmes d'une radio publique spécialisée en diffusion de musiques baroque et classique n'est altérée par aucun effet de congestion. À l'inverse, la satisfaction éprouvée en flânant dans les allées d'un parc municipal ou en s'allongeant sur le sable chaud d'une plage publique pourra être grandement dégradée si le parc ou la plage sont envahis par une foule conséquente et/ou un groupe d'individus bruyants et indisciplinés. Notons cependant que le caractère immatériel de la prestation ne garantit pas l'absence d'effets d'encombrements : l'accès aux programmes en *replay* d'une télévision publique peut être de mauvaise qualité ou fréquemment interrompu si le nombre des téléspectateurs simultanément connectés est trop grand au regard des capacités du serveur. On remarquera enfin que les effets d'encombrements s'apparentent à des externalités : c'est la présence ou l'action de tiers qui dégradent la satisfaction éprouvée par l'agent économique considéré.

Les **biens publics purs** sont des biens et services qui réunissent les spécificités d'impossibilité d'exclusion par l'usage, d'obligation d'usage et d'absence d'effets d'encombrements.

On peut proposer des classifications des biens publics selon qu'ils réunissent seulement deux, ou même une seule, des trois spécificités ci-dessus listées. L'un des enjeux est de différencier les biens publics selon la manière dont leur fourniture pourra être assurée. En effet, puisque la puissance publique va devoir participer à la fourniture du bien public (ou en assurer pleinement la fourniture), il est opportun de savoir si sa fourniture sera financée par des impôts et taxes, des recettes « aux guichets » ou les deux. Dans le cas d'une piscine municipale ou de transports urbains, les recettes permettant de financer la fourniture du bien public proviennent en général des deux sources. En revanche, dans le cas de la collecte et du traitement de déchets, le financement des prestations est, en général, assuré par les seuls impôts et taxes. Enfin, une dernière question est essentielle : le caractère du bien public est-il local ou supra local ? Il faut en effet savoir à quel échelon de collectivité territoriale revient la responsabilité de fourniture : ainsi, les transports en communs, le réseau d'évacuation des eaux usées et la collecte des déchets ont vocation à être fournis ou assurés par la communauté de communes dans une aire urbaine ; un service de protection de la faune et de la flore peut être attaché à un massif montagneux particulier ; des services de surveillance sismique ou de prévision météorologique sont possiblement attachés à un continent tout entier…

Pourquoi la fourniture des biens collectifs nécessite-t-elle l'intervention de l'État ou d'une collectivité territoriale ? La théorie microéconomique répond à cette question en indiquant que si on laisse chaque agent décider librement de la contribution qu'il souhaite apporter pour que ce type de biens soit fourni, le montant total récolté est insuffisant pour que ces biens puissent être produits et/ou fournis en quantité satisfaisante (bien que chacun soit désireux de voir ces biens mis à disposition des consommateurs). Comment expliquer ce phénomène ? Par les comportements dits de « **passagers clandestins** ». Supposons qu'il y ait 300 000 habitants dans une ville et que le coût annuel de l'éclairage public soit de 299 999 euros. Cha-

cun se dit que la contribution de tous les autres habitants à hauteur de 1 € va suffire à financer ce service et qu'il n'est donc pas nécessaire de contribuer soi-même. Comme chacun fait le même raisonnement, les recettes de nature à financer la fourniture d'éclairage public, sur la base de contributions volontaires, sont donc nulles et il y a une « sous-fourniture » du bien. On peut affiner cette analyse en caractérisant l'allocation spontanée des ressources de l'économie lorsque chacun est amené à financer le bien public sur la base d'une contribution volontaire et en comparant avec ce qu'un « planificateur » soucieux de la fourniture suffisante du bien public déciderait.

Reprenons le travail effectué dans le cadre du chapitre 5 ou deux consommateurs, *A* et *B*, cherchaient à améliorer conjointement leur sort en échangeant deux biens indicés 1 et 2. Nous étions parvenus à la conclusion que ceux-ci parvenaient à une allocation d'équilibre telles que $TmS_{1/2}{}^{A} = TmS_{1/2}{}^{B}$: le taux marginal de substitution entre les biens 1 et 2 est identique pour les individus A et B. En d'autres termes, quand les agents choisissent sans contrainte les quantités de deux biens qu'ils vont consommer (de manière dite « décentralisée »), ils parviennent à une allocation qui égalise leurs *TmS*. Dans ce même chapitre 5, nous avions mentionné que le cadre présenté pouvait être élargi en tenant compte de la production des biens. Si l'on élargit en effet cette analyse en tenant compte de la production de l'un des biens, il est alors possible et opportun de compléter la caractérisation de l'équilibre décentralisé. Supposons désormais que le bien 1 est un bien public dont on modélise la production et que le bien 2 est un bien privé (dont on ne modélise pas la production). La caractérisation complète de l'**allocation d'équilibre décentralisé** de l'économie avec production de bien public est la suivante :

$$|TmS_{1/2}{}^{A}| = |TmS_{1/2}{}^{B}| = Cm(x_1)$$

où $Cm(x_1)$ désigne le coût marginal de production du bien public.

Cette double égalité nous indique qu'à l'équilibre, les taux marginaux de substitution entre le bien public et le bien privé doivent être :

- identiques pour les deux agents et,
- égaux au coût marginal de production du bien public.

Pour caractériser la décision optimale d'un **planificateur** s'efforçant de **maximiser le *welfare*** obtenu dans cette économie, on obtient l'égalité suivante :

$$|TmS_{1/2}{}^{A}| + |TmS_{1/2}{}^{B}| = Cm(x_1)$$

Cette condition est une écriture simplifiée de la condition dite de **Bowen-Lindahl-Samuelson** (BLS).

En comparant la condition d'équilibre décentralisé et la condition de maximisation du *welfare*, on constate que le coût marginal de production du bien public doit être, à l'optimum du planificateur, supérieur à ce qu'il est à l'équilibre décentralisé : en effet la somme des valeurs absolues des TmS doit être supérieure à la valeur absolue d'un quelconque des TmS. Or, puisque le coût marginal de production du bien public est supposé croissant, cela signifie que la production de bien public à l'équilibre décentralisé est inférieure à ce qu'elle serait à l'optimum du planificateur. Il y a donc sous fourniture du bien public sans intervention de la puissance publique. C'est pourquoi, la restauration de l'efficacité économique passe par l'intervention publique : l'État et/ou les collectivités publiques font payer des impôts et taxes aux consommateurs pour obtenir des recettes afin de financer la fourniture des biens publics en quantité satisfaisante.

Pour aller plus loin

La condition BLS

Soit un planificateur dont la fonction objectif est $W = \sum_{i=1}^{n} \alpha_i u_i(x_i; g)$ où les $u_i(x_i ; g)$ désignent les fonctions d'utilité des n agents de l'économie, toutes définies sur la quantité consommée d'un bien privé, notée x, et la quantité consommée d'un bien public, notée g. Les α_i désignent les valeurs sociales des différents individus, c'est-à-dire le poids accordé par le planificateur à la satisfaction de chaque individu. On suppose que la production du bien public se fait en supportant un coût total $C(g)$.

Si l'on maximise le bien-être social (le *welfare*) sous contrainte technologique, on obtient la condition d'optimalité suivante :

$$\alpha_1 \frac{\frac{\partial u_1}{\partial g}}{\frac{\partial u_1}{\partial x_1}} + \alpha_2 \frac{\frac{\partial u_2}{\partial g}}{\frac{\partial u_2}{\partial x_2}} + \ldots + \alpha_n \frac{\frac{\partial u_n}{\partial g}}{\frac{\partial u_n}{\partial x_n}} = \frac{\partial C(g)}{\partial g}$$

$$\Leftrightarrow \alpha_1 |TmS_{g/x}{}^1| + \alpha_2 |TmS_{g/x}{}^2| + \ldots + \alpha_n |TmS_{g/x}{}^n| = Cm(g) \qquad \textit{Condition BLS}$$

12 La prise en compte des asymétries d'information

Mots-clés

Modèle principal-agent, anti-sélection, contrats séparateurs, alea moral

Pourquoi n'existe-t-il pas d'assurance privée contre la pauvreté ou la misère ? À une époque où les compagnies d'assurance rivalisent d'ingéniosité pour proposer de nouvelles couvertures assurantielles, on peut s'étonner de l'absence de proposition de contrats d'assurance contre le risque de devenir très pauvre. *A priori*, il devrait être possible de calculer les probabilités des différents événements négatifs pouvant survenir ainsi que le montant des indemnités qu'il faudrait verser pour que l'assuré sorte de la grande pauvreté ; dès lors il semble théoriquement possible de calculer le montant des primes d'assurance à acquitter pour être couvert contre un tel risque. Mais seuls les individus possédant un fort risque de devenir très pauvres seraient désireux de souscrire un tel contrat, alors que les individus présentant un faible risque n'auraient aucune intention d'y souscrire. Un tel contrat, caractérisé par une prime d'assurance d'un montant unique et une indemnité en cas de « sinistre » (la survenance d'une situation de grande précarité) elle aussi d'un montant unique, serait trop cher au regard du risque encouru pour les faibles risques et très attractif au regard du risque encouru pour les risques élevés. Les faibles risques (des agents diplômés, propriétaires de leur logement, en bonne santé, peu endettés, ayant une bonne insertion professionnelle et une situation familiale stable) seraient donc absents du marché et les recettes de la compagnie d'assurance, sous forme de primes acquittées par les assurés, diminuées d'autant. Seuls les risques élevés (des agents sans qualification, non propriétaires



d'un logement, en mauvaise santé, surendettés, en situations professionnelle et familiale précaires) seraient présents et le niveau de réalisation des « sinistres » serait beaucoup trop grand pour que les primes collectées suffisent à couvrir les indemnités d'assurance contre la grande pauvreté que la compagnie d'assurance devrait verser. Si une entreprise d'assurance privée proposait un tel contrat, elle croulerait rapidement sous des pertes abyssales. Aucune entreprise privée ne peut donc assurer les agents contre la misère, un tel marché d'assurance ne peut donc exister. Pour décrire le phénomène détaillé ci-dessous, on parle d'**anti-sélection** : les termes du contrat d'assurance que proposerait une entreprise d'assurance privée dissuadent les « bons » risques de souscrire et encouragent les « mauvais » risques. Le résultat est que seuls les « mauvais » risques affluent : la compagnie d'assurance a recueilli les « pires » clients envisageables. Dans le cas présent, seule l'intervention de la puissance publique peut apporter une réponse à cette « malédiction » de l'anti-sélection : l'État crée une assurance obligatoire (schématiquement dénommée « assurance sociale ») et peut ainsi mettre en place des prestations au bénéfice de ceux qui connaissent des situations difficiles (assurance chômage, assurance-maladie, assurance vieillesse, revenus de solidarité, etc.). En réalité, il n'est plus question de primes et d'indemnités d'assurance (privée) mais de cotisations et de prestations (sociales). Le phénomène d'anti-sélection est un des phénomènes que les économistes analysent sous la rubrique des asymétries d'information. Nous allons, dans ce chapitre, analyser les situations où deux partenaires d'une relation économique (comme un assureur et un assuré) ne possèdent pas le même niveau d'information sur une caractéristique de l'un des partenaires. Cette situation dite d'information imparfaite conduit à des dysfonctionnements du marché considéré voire même à la disparition totale du marché. Il existe, cependant, des dispositifs permettant de révéler l'information inconnue et de corriger, au moins partiellement, la perte d'efficacité sociale consécutive à l'existence des asymétries informationnelles. L'élaboration de contrats révélateurs est l'un des aspects les plus stimulants de cette branche récente de la microéconomie, mais d'autres dispositifs existent, comme les processus de certification qui informent le consommateur sur la qualité réelle des biens et services vendus.

Une autre forme d'asymétrie d'information peut également perturber le fonctionnement des marchés : c'est l'**aléa moral**. L'aléa moral est une situation où deux partenaires d'une relation économique ne possèdent pas le même niveau d'information sur l'action ou l'effort entrepris par l'un des partenaires. Par exemple, lorsque des parents confient la garde de leur enfant à une assistante maternelle, une crèche ou une garderie scolaire, il leur sera difficile de savoir si leur enfant aura été stimulé par des activités reposant sur la lecture et des jeux éducatifs ou, à l'inverse, tiré vers le néant intellectuel en étant prostré devant un écran de télévision, d'ordinateur ou de console de jeux toute la journée (notons au passage que la probabilité d'être laissé devant un écran de télévision est plus forte auprès d'une assistante maternelle qui travaille seule chez elle, que dans une crèche où collaborent plusieurs assistantes et, qui, consciemment ou inconsciemment, exercent une forme de contrôle les unes sur les autres). Quand un employé municipal se voit confier la mission de ramasser, tout au long de la journée, les détritus jetés par les passants sur un lieu touristique, son employeur saura difficilement s'il a effectivement œuvré en permanence à la propreté du lieu ou s'il s'est contenté d'attendre le passage des gros engins de nettoyage à la fin de la journée. Les actionnaires d'une grande entreprise auront, eux aussi, une certaine difficulté à observer l'effort fourni par le *manager* pour améliorer (ou maintenir) la profitabilité de l'entreprise : ce dernier saura mieux que les actionnaires s'il a travaillé avec détermination et efficacité ou, au contraire, s'il se sera comporté en pacha oisif. Evidemment, il peut exister des dispositifs de contrôle, précisément mis en place pour éviter les dysfonctionnements que ces situations engendrent (contrôles aléatoires, caméras de surveillance, etc.) mais surtout il y a des formes particulières de contrats, qui régissent la relation entre les deux partenaires de la relation économique, qui incitent à l'effort. C'est une des solutions que l'on peut objectiver par la théorie de l'information et des incitations, discipline, au sein de la microéconomie, qui se penche sur les conséquences de l'existence d'asymétries d'information et les solutions qui peuvent y être apportées.

1 Théorie de l'agence et modèle principal-agent

Dans les chapitres précédents, nous avons explicitement ou implicitement supposé que l'information était parfaite : chaque acteur des relations économiques était réputé connaître parfaitement les caractéristiques de ses clients, partenaires ou rivaux. Il va de soi que, dans la réalité, une grande part de ces caractéristiques est mal connue ou inconnue. Ceci n'est pas sans conséquence sur le fonctionnement des marchés et sur la vraisemblance d'un cheminement spontané vers l'efficacité sociale. Pire encore, le marché peut même disparaître en raison de la présence d'asymétries d'informations : dans ce cas, le *welfare* relatif à l'existence du marché s'évanouit complètement. Depuis les années 1970, la microéconomie s'est enrichie de nouveaux développements visant à analyser les relations économiques en contexte d'information imparfaite. La modélisation de ces relations repose sur l'hypothèse qu'il existe un agent convenablement informé et un agent imparfaitement informé. Dans cet ouvrage, nous allons nous limiter aux cas où l'agent imparfaitement informé est celui qui dessine les contours du contrat qui le liera avec l'agent bien informé. Par exemple, un employeur propose un contrat de travail à un salarié sans connaître parfaitement ses aptitudes et/ou son tempérament plus ou moins « travailleur ». Le pouvoir de proposition des termes du contrat est ce qui caractérise le partenaire désigné sous l'appellation de « principal » ; l'autre partenaire est appelé « agent ». Cette terminologie, bien que très utilisée, est à l'origine de confusions. En effet, il peut aussi exister des cas où c'est le « principal » qui est convenablement informé. Or, l'élément clé de l'analyse est le caractère parfaitement ou imparfaitement informé de chacun des partenaires.

La **théorie de l'agence** désigne les travaux qui analysent la relation économique entre un partenaire convenablement informé et un partenaire imparfaitement informé. Le partenaire qui propose les termes du contrat est appelé « **principal** », l'autre partenaire est appelé « **agent** ».

On distingue deux types d'asymétries d'information : le cas où l'information imparfaitement connue concerne une caractéristique intrinsèque de l'individu parfaitement informé (qu'il est supposé connaître lui-même) et le cas où l'information imparfaitement connue concerne une action entreprise par l'individu parfaitement informé. Dans le premier cas, il s'agira d'une situation d'anti-sélection, dans le second cas d'une situation d'aléa moral. Le cas de l'anti-sélection sera traité dans la section suivante. Nous analyserons l'aléa moral dans une troisième section.

2 L'anti-sélection

Le terme anti-sélection est la traduction de l'expression anglaise *selection adverse*. Cette expression évoque plus les conséquences d'une situation informationnelle particulière que sa nature ou son origine : en effet, c'est parce que les caractéristiques des agents sont mal connues de celui qui leur propose un contrat (d'assurance, de travail, etc.), que le contrat qui sera proposé ne sera « signé » que par les agents dotés de mauvaises caractéristiques (risque élevé, aptitudes faibles, ...). Pour cette raison, on parlera d'un phénomène d'anti-sélection : la proposition faite n'a attiré que ceux que le partenaire à l'origine de la relation économique ne voulait pas « sélectionner » !

L'**anti-sélection** est une configuration informationnelle où l'information imparfaitement connue concerne une caractéristique intrinsèque de l'individu parfaitement informé et/ou du bien ou service qu'il propose.

Les exemples de telles configurations informationnelles ne manquent pas. L'exemple emblématique est celui du « marché » des voitures d'occasion puisque c'est à partir de cet exemple que George Akerlof a initié les travaux sur les asymétries d'information. Lorsqu'un automobiliste achète une voiture d'occasion, il ne connaît pas parfaitement l'état du véhicule qu'il achète. Le vendeur, à défaut de connaître parfaitement l'état de son véhicule, sait s'il met en vente une voiture parfaitement « menée » et entretenue

ou, à l'inverse, une quasi-épave prête à agoniser au bord de la route. Un autre exemple majeur concerne aussi les automobilistes : c'est celui de l'assurance que souscrivent les conducteurs pour couvrir les dégâts occasionnés par les accidents de la route ou les dégradations subies par leurs véhicules. L'assureur ne sait pas parfaitement si le souscripteur est un conducteur habile, concentré et prudent ou au contraire maladroit, distrait et imprudent. Il connaît mal le nombre de kilomètres parcourus par l'automobiliste et le temps que son véhicule passera garé dans un garage fermé ou, à l'opposé, sur des parkings régulièrement visités par des désoeuvrés pyromanes. Sans lien avec l'exemple précédent, on peut aussi indiquer qu'un employeur ne connaîtra jamais parfaitement les aptitudes réelles d'un candidat à l'embauche, en particulier lorsque l'emploi proposé ne requiert aucune qualification particulière susceptible d'être attestée par un diplôme ou un certificat. Enfin, lorsqu'un banquier accorde un crédit à un emprunteur, il ignore largement si celui-ci est une « pure » cigale telle qu'évoquée dans le chapitre 3 ou un consommateur aux préférences intertemporelles strictement convexes.

La présence de l'asymétrie informationnelle peut conduire à la disparition (ou à la non apparition) du marché. Le mécanisme de disparition progressive du marché est simple à comprendre. Si nous reprenons l'exemple des automobiles d'occasion, nous devinons que, très probablement, face à l'incertitude sur la qualité des véhicules mis en vente, l'acheteur sera disposé à payer un prix plus faible que le prix réclamé par un vendeur de véhicule de bonne qualité. Même en l'absence d'aversion pour le risque, l'acheteur qui penserait qu'il a une chance sur deux d'acheter un « bon » véhicule d'occasion (pour lequel il est prêt à payer, disons 5 000 €) et une chance sur deux d'acheter un « mauvais » véhicule d'occasion (pour lequel il est prêt à payer, au maximum, 1 000 €), ne voudra pas dépenser plus de ½ × (5 000 + 1 000) = 3 000 €. Si, en face, le vendeur d'une « bonne » occasion considère que son véhicule ne vaut pas moins de 4 000 €, il décidera de retirer de la vente son véhicule. Par ce type de raisonnements, il apparaît clairement que, progressivement, tous les véhicules de bonne qualité vont disparaître du marché puisque la disponibilité minimale à recevoir des vendeurs est supérieure à la disponibilité maximale à payer des acheteurs. Il en résulte l'existence d'un marché uniquement composé de vieux tacots rêvant de Provence,

mais voués à mourir à Fontainebleau ! De même, dans un monde où n'existerait pas de salaire minimum, face à l'incertitude sur la qualité (et donc la productivité) des travailleurs, un employeur serait amené, à l'issue d'un calcul rationnel, à proposer un unique salaire trop faible pour attirer les « bons » salariés. Ne maintiendraient alors leurs candidatures que les inaptes et les incompétents … Notons, bien entendu, que la présence d'un salaire minimum n'élimine pas l'anti-sélection (elle peut même en aggraver les effets si le salaire minimum alourdit le coût unitaire de la firme sans être, pour autant, suffisant pour attirer les « bons » travailleurs).

Comment répondre à la présence d'anti-sélection et aux dysfonctionnements qu'elle entraîne ? Nous nous concentrerons sur deux types de réponses : la certification et la mise en place d'un menu de contrats. Les deux réponses ne sont pas de même nature. Dans certaines circonstances, elles ne peuvent, ni l'une, ni l'autre, n'être suffisantes pour lutter contre les dysfonctionnements du marché. Néanmoins, elles constituent des démarches opportunes et parfois très astucieuses pour contrer les effets indésirables de la présence d'imperfections informationnelles.

2.1 La certification

Les processus de certifications (qui sont une réponse à l'anti-sélection) existent préalablement à la découverte, par la théorie économique, des phénomènes qu'ils permettent de corriger. C'est une réponse appropriée des producteurs de biens de bonne qualité ou des travailleurs compétents à la malédiction de l'éviction de leurs produits ou de leurs talents des marchés où ils auraient vocation à jouer le rôle principal. Pour un travailleur, le principal processus de certification est le diplôme au sens large (brevet ou certificat d'aptitude professionnelle, test de maîtrise d'une langue, attestation d'un précédent employeur, etc.) : en obtenant un diplôme, le travailleur envoie un signal sur le niveau de ses aptitudes, connaissances et/ou compétences. Le travailleur apte peut ainsi être distingué du travailleur inapte par le principal. Remarquons néanmoins, que, face à la multiplicité des diplômes, les organismes qui les délivrent (universités, écoles, instituts, etc.) cherchent eux-mêmes une certification supérieure (certification par l'État, par une commission des titres, par un réseau d'instituts, etc.). Par-

fois même, ceux qui certifient les organismes qui délivrent des diplômes cherchent encore une certification plus haute : c'est ainsi que l'on peut comprendre les rapprochements ou les mises en réseaux dans le cadre de programmes ou d'accords internationaux... Dans le cas des voitures d'occasion, un premier élément de certification est le contrôle technique. Il s'agit d'une certification imposée par la puissance publique pour améliorer la sécurité sur les routes, mais aussi pour apporter des garanties aux acheteurs. Une autre manière de lutter contre l'anti-sélection est d'acheter son véhicule d'occasion chez un professionnel. On peut supposer que celui-ci ne remet sur les routes que des véhicules aptes à rouler et pour lesquels les travaux essentiels ont été accomplis. Si nous reprenons l'exemple présenté plus haut, le détenteur d'un véhicule de bonne qualité – qui ne souhaite pas le céder à moins de 4 000 € – le vendra à un professionnel de l'occasion pour 4 200 €. Le professionnel le revendra lui-même 4 800 € à un acheteur dont la disponibilité maximale à payer est, nous l'avons indiqué, de 5 000 €. Chacun gagne en *welfare* et les véhicules de bonne qualité ne sont plus évincés du marché de l'occasion. La certification a eu un coût : 600 € que le vendeur n'aura pas obtenu de l'acheteur alors que celui-ci aurait consenti à les lui donner en situation d'information parfaite. La certification par un vendeur de véhicule d'occasion n'est crédible que si le vendeur est lui-même réputé pour la qualité de son travail de réparateur, qui atteste de ses compétences et donc de la légitimité de sa certification (la certification par un vendeur de véhicules d'occasion qui n'exerce aucune activité de réparation ou d'entretien n'a, qu'on se le dise, aucune valeur...). Les appellations contrôlées et les labels sont, eux aussi, des dispositifs visant à lutter contre l'anti-sélection. Qu'ils soient délivrés par des autorités indépendantes ou sous contrôle d'un organisme gouvernemental, ils permettent aux firmes qui les obtiennent de signaler la qualité de leurs produits : garantie de provenance, qualité sanitaire de la filière, respect de bonnes pratiques dans l'élaboration du produit, qualité nutritive ou gustative minimale...

2.2 La mise en place d'un menu de contrats

La partie la plus stimulante des travaux sur l'anti-sélection se trouve dans la recherche des mécanismes d'incitations (*mechanism design*) qui conduisent les agents à révéler leur type. Plus précisément, il s'agit de

construire un menu de contrats subtilement différenciés, entre lesquels les agents auront le libre choix, et tels que les agents de chaque type choisissent le contrat qui leur est dédié. Les contrats sont alors révélateurs et il est alors possible de ne pas faire disparaître les « bons » agents du marché (sans toutefois que leur *welfare* ne soit aussi élevé que ce qu'il aurait été en situation d'information parfaite).

Un exemple très simple de ces subtils menus de contrats concerne l'assurance. Supposons que, pour une catégorie de sinistre particulière, existent deux types d'agents : les « gros » risques et les « petits » risques. Par exemple, en matière de conduite automobile, certains agents sont maladroits, sujets à la distraction et très impulsifs tandis que d'autres sont très précis dans leur conduite, toujours concentrés et d'un grand calme. Supposons que l'assureur (le principal), ne connaisse pas le « type » (gros risque ou petit risque) de chaque agent (il ne dispose, en particulier d'aucune information sur le sexe, l'âge, le kilométrage annuel et l'historique d'accidents de l'agent). S'il propose un unique contrat d'assurance correspondant à un risque moyen, il verra les petits risques fuir le contrat proposé car la prime d'assurance sera trop élevée au regard de l'indemnité et/ou de la probabilité de sinistre les concernant. On retrouve les arguments développés en introduction à propos de « l'introuvable assurance contre la misère ». Cette fois-ci pourtant, on peut envisager, sous certaines conditions, l'émergence d'un dispositif susceptible de convenir aux gros risques comme aux petits risques. Ce dispositif nécessite l'existence d'une **franchise**.

Dans un contrat d'assurance, la franchise (déductible) est un montant qui sera retranché à la perte subie pour déterminer le niveau de l'indemnité (en cas de réalisation du sinistre). Par exemple, si, dans un contrat d'assurance-automobile, est prévue une franchise de 300 €, le montant de l'indemnité versée à l'assuré suite à un sinistre qui aura entraîné 1 500 € de réparations est 1 500 – 300 = 1 200 €. Lorsqu'un contrat d'assurance est assorti d'une franchise, l'assuré se désole de ne pas pleinement être indemnisé, mais il se réjouit de payer une prime d'assurance moins élevée que ce qu'elle aurait été en l'absence de franchise.

Plus en détail, la théorie économique montre que les agents sont supposés préférer les contrats avec franchise (plutôt que des contrats sans franchise)

au terme d'un arbitrage entre l'effet « prime » d'assurance et l'effet « couverture » du risque… Mais revenons à nos deux catégories d'agents différenciés par le niveau de risque qu'ils « portent en eux ». Ce que l'assureur a intérêt à leur proposer est de choisir librement entre deux contrats :

- un contrat *A* avec une forte franchise et une faible prime d'assurance ;
- un contrat *B* avec une faible franchise et une forte prime d'assurance.

Si les contrats sont astucieusement construits, tous les « gros risques » vont choisir le contrat *B* et tous les « petits risques » vont choisir le contrat *A*. En effet, un agent présentant un gros risque de sinistre sait qu'il est fort probable qu'il soit impliqué dans un accident matériel dans le courant de l'année. S'il est assujetti à une forte franchise, il sait que cet accident aura des conséquences fort coûteuses. Il préfère donc payer une prime d'assurance relativement élevée mais qui lui assure une très bonne couverture. Du côté des agents présentant un petit risque de sinistre, le raisonnement est inverse. Ils vont tous choisir le contrat *A* car ils savent qu'il est peu probable qu'ils soient impliqués dans un accident au cours de la période. Certes, la franchise à supporter est élevée, mais il y a très peu de chance qu'ils aient à la subir. Ils privilégient donc le contrat dont la prime est très faible (en d'autres termes, tandis que les gros risques raisonnent comme s'ils allaient avoir un accident dans le courant de l'année, les petits risques raisonnent comme s'ils n'allaient pas en avoir). Le résultat semble presque magique : chacun a spontanément choisi le contrat qui lui était dédié. Le menu de contrats proposé a été **révélateur** : les gros risques ont choisi le contrat *B*, les petits risques ont choisi le contrat *A*.

La possibilité de mettre en place un menu de contrats qui sera un mécanisme révélateur n'est pas garantie. Il faut que certaines propriétés soient vérifiées, en particulier quant aux proportions relatives de gros et de petits risques (de manière un peu contre intuitive, on montre qu'il est nécessaire que la proportion de petits risques ne soit pas trop forte !). En outre, le contrat proposé aux petits risques est moins avantageux que ce qu'il aurait été en l'absence d'asymétries d'information. De plus, la rivalité concurrentielle entre les différents assureurs peut remettre en cause la capacité de chacun à proposer un menu de contrats révélateurs. Néanmoins, nous avons esquissé la notion d'**équilibre séparateur** qui est susceptible d'émer-

ger lorsqu'un astucieux menu de contrats est proposé en réponse à la présence d'anti-sélection.

Dans certains cas, les dysfonctionnements liés à l'anti-sélection ne peuvent être jugulés par aucun dispositif. L'intervention de la puissance publique est alors nécessaire pour, à défaut de restaurer le marché disparu ou fonctionnant très mal, le remplacer par un dispositif contraignant voulu par le législateur : c'est ainsi que l'on peut légitimer, dans le cadre de la théorie économique, les dispositifs publics d'assurance-maladie, vieillesse, chômage, etc. financés par les cotisations de l'ensemble des citoyens.

3 L'aléa moral

L'aléa moral (*moral hazard en anglais*) désigne les situations où l'effort entrepris par un agent est inobservable (ou imparfaitement observable) par l'autre partenaire de la relation économique. C'est le cas pour des parents qui ont confié la garde de leur enfant à une assistante maternelle et qui ne peuvent qu'imparfaitement observer l'effort entrepris par cette dernière ; c'est le cas pour un supérieur hiérarchique dans une entreprise qui n'observe qu'imparfaitement la détermination et le zèle d'un subalterne ; c'est aussi le cas pour un automobiliste qui ne peut savoir si le garagiste, à qui il a confié son véhicule, a en effet consacré quatre heures aux diverses réparations (comme il est indiqué sur la facture...). L'effort entrepris par l'agent peut être un effort productif ou un effort de type « je roule plus prudemment », qui affecte (positivement) le sort du principal qu'est ici l'assureur.

L'aléa moral est une configuration informationnelle où l'information imparfaitement connue concerne l'intensité d'un effort ou d'une action entreprise par l'individu parfaitement informé.

Pour bien comprendre le sens du terme aléa moral, il faut prendre la mesure des éventuelles conséquences de l'asymétrie informationnelle. En effet, dans le contexte d'un contrat d'assurance, c'est plutôt la crainte de ce que l'agent, se sachant assuré, agisse de manière imprudente et/ou inconséquente qui

justifie le choix de la terminologie « d'aléa moral ». Un armateur, se sachant assuré, pourrait choisir de ne faire naviguer que de très vieux navires mal entretenus qui, par mauvaise fortune, finiraient par s'échouer sur de quelconques côtes. Une banque, se sachant « assurée » par l'État (enclin à garantir la sécurité des dépôts des épargnants) pourrait se hasarder dans des opérations financières très juteuses mais excessivement risquées. Un chômeur, se sachant couvert par une assurance chômage, pourrait ne pas fournir un effort très intensif de recherche d'emploi... Cette dimension du problème posé par l'existence d'aléa moral nous permet d'évoquer quelques pistes de dispositifs permettant de le juguler : l'assureur, privé ou public, peut prévoir, dans son contrat ou dans la réglementation, des dispositions contraignantes. Ainsi l'armateur devra accepter des contrôles sévères de ses navires (par la compagnie d'assurance et par une autorité maritime préoccupée par les dangers que font courir les navires « poubelles ») ; ainsi les banques devront respecter des ratios prudentiels (Bâle II, Bâle III, ...) garantissant leur solidité financière ; ainsi le chômeur indemnisé devra subir des contrôles du caractère effectif de sa recherche d'emploi... Mieux encore, la théorie économique part à la recherche de formes de contrats incitant l'agent à fournir un effort optimal.

Comme son nom l'indique, le **contrat de fermage** tire son origine de pratiques issues du secteur agricole. Le contrat de fermage est une forme très ancienne de contrats entre un propriétaire terrien et l'exploitant de la terre considérée, en vertu duquel l'exploitant verse un montant fixe, appelé loyer, au propriétaire et jouit pleinement, en contrepartie, du produit de la terre. Ceci conduit, certes, à ce que l'exploitant supporte tout le risque de mauvaise récolte, mais le contrat est pleinement incitatif pour l'agent. Pour le comprendre, il faut comparer le contrat de fermage à ce qui se passerait si l'agent était un salarié. Assuré de percevoir un salaire fixé à l'avance, l'exploitant fournirait un effort minimal qui conduirait à des récoltes très faibles (peut-être même nulles) et ceci, indépendamment des aléas climatiques. On peut montrer de manière assez simple qu'à l'inverse, sous certaines conditions (en particulier l'absence d'aversion pour le risque de l'agent), un contrat de fermage conduit à un niveau d'effort optimal, équivalent à celui qui serait fourni en l'absence d'asymétries d'information.

Le contrat de fermage est très répandu mais souvent identifié par une autre dénomination : la franchise. De nombreux magasins fonctionnent sous contrat de franchise avec une « enseigne » nationalement ou internationalement connue. Les droits de propriété détenus par le principal ne concernent plus le terrain (ou les murs du magasin) mais la marque (ou son utilisation et ses identifiants marketing). Souvent, le contrat de franchise offre le droit (ou entraîne l'obligation) de distribuer certains biens et services (produits par le principal ou par un tiers). L'élément essentiel est que le franchisé est le *residual claimant* (le créancier résiduel), c'est-à-dire un agent qui, après avoir acquitté le montant du « loyer » (appelée aussi… franchise), conserve l'intégralité des gains d'exploitation réalisés. Ce type de contrats, incitatif à l'effort, est très répandu dans la distribution. Il est aussi considéré, par la théorie économique, comme une solution alternative à l'intégration verticale : plutôt que d'acquérir les structures de distribution (les « points de vente ») et de devoir mettre en place des dispositifs coûteux de contrôle de leur gestion, il convient de laisser à des indépendants, liés par un contrat de franchise, le soin de commercialiser les produits d'une firme.

L'analyse des asymétries d'information est un des champs les plus actifs et les plus stimulants de la théorie économique moderne. Pour sa gouverne, le lecteur aura appris qu'il ne faut pas acheter son véhicule auprès d'un vendeur de véhicule d'occasion qui n'est pas aussi un réparateur (si possible jouissant d'une réputation de professionnel compétent) et que lorsqu'il choisit le contrat d'assurance que lui propose un assureur, il doit privilégier le contrat avec la plus forte franchise pour se signaler comme un « bon » risque…

Glossaire

Aléa moral / ***Moral hazard***
Configuration informationnelle où l'information imparfaitement connue concerne l'intensité d'un effort ou d'une action entreprise par l'individu parfaitement informé.

Anti-sélection / ***Selection adverse***
Configuration informationnelle où l'information imparfaitement connue concerne une caractéristique intrinsèque de l'individu parfaitement informé et/ou du bien ou service qu'il propose.

Arbitrage / ***Trade-off***
Décision issue de la comparaison entre différentes opportunités (ou entre différents « maux »). En économie, l'arbitrage optimal est atteint lorsque sont égalisées des grandeurs marginales.

Bien contingent / ***Contingent good***
Bien consommé pour une issue particulière d'une épreuve aléatoire, c'est-à-dire, pour un *état de la nature* particulier.

Biens publics / ***Public goods***
Biens et services qui réunissent tout ou partie des spécificités suivantes : impossibilité d'exclusion par l'usage, obligation d'usage, absence d'effets d'encombrements.

Boîte de Pareto-Edgeworth / ***Edgeworth box***
La boîte de Pareto-Edgeworth d'une économie est un rectangle figurant toutes les répartitions possibles des consommations de deux biens entre deux agents économiques. La longueur et la largeur de cette boîte correspondent aux quantités totales présentes des deux biens dans l'économie.

Cartel / ***Cartel***
Sous ensemble de firmes ayant décidé de s'entendre sur le prix et les quantités à produire d'un même bien ou service. Synonyme d'entente ou de collusion.

Concurrence pure et parfaite / ***Perfect competition***
Configuration de marché pour laquelle la production d'un bien ou service est assurée par une multitude de firmes de très petite taille produisant toutes un bien identique.

Condition équi-marginale / ***Equi-marginal condition***
Condition caractérisant la décision optimale d'un agent. Dans le cas de la décision du consommateur, la condition équi-marginale dispose que les surcroîts de satisfaction par euro dépensé, obtenus grâce aux dernières unités consommées de chaque bien, doivent être égales.

Contrainte budgétaire / ***Budget constraint***
Inégalité qui caractérise l'ensemble des paniers de biens que le consommateur est susceptible d'acquérir étant donné son revenu et les prix auxquels les biens sont vendus.

Courbe des contrats / ***Contract curve***
Courbe des optima de Pareto.

Courbe d'Engel / ***Engel curve***
Courbe représentant la consommation optimale d'un bien par un consommateur en fonction de son revenu.

Courbe d'indifférence / ***Indifference curve***
Ensemble des paniers de biens dont la consommation apporte un même niveau d'utilité à un consommateur particulier.

Coût marginal / ***Marginal cost***
Surcroît de coût lié à la production d'une unité supplémentaire d'output.

Coût moyen / ***Average cost***
Prix de revient d'une unité d'output produite.

Coût total de production / ***Total cost function***
Valeur minimale de la dépense à consentir pour acquérir les inputs utilisés pour la production d'une certaine quantité d'output.

Discrimination tarifaire / *Price discrimination*
Situation où une firme peut vendre son output à des prix différents auprès de différents consommateurs.

Duopole de Cournot / *Cournot duopoly*
Configuration de marché dans laquelle deux firmes en positions symétriques se font concurrence en quantité.

Duopole de Stackelberg / *Stackelberg duopoly*
Configuration de marché dans laquelle deux firmes en positions asymétriques (un leader et un follower) se font concurrence en quantité.

Effet externe / *Externality*
Phénomène dont l'origine est une action d'un agent économique affectant la satisfaction ou l'ensemble de choix d'un ou plusieurs autres sans que les éventuelles variations de prix conséquentes à cette action ne véhiculent convenablement l'information sur l'état de l'allocation des ressources dans l'économie.

Effet substitution / *Substitution effect*
Effet, sur la demande des biens, de la modification des prix relatifs des biens et qui conduit le consommateur à consommer moins des biens devenus relativement plus chers et plus des biens devenus relativement moins chers.

Effet revenu / *Income effect*
Effet, sur la demande des biens, de la variation de pouvoir d'achat induite par la variation du revenu relatif et qui pousse le consommateur à consommer moins de chacun des biens lorsque le revenu relatif diminue, et plus de chacun des biens lorsque le revenu relatif augmente.

Élasticité de la demande / *Demand elasticity*
Mesure de sensibilité de la demande d'un bien aux variations d'une des grandeurs dont elle dépend (prix du bien considéré, prix d'un autre bien, revenu du consommateur).

Équilibre / *Equilibrium*
L'équilibre est une allocation des ressources telle que, si les agents l'ont atteinte, ils ne s'en écarteront pas.

Équilibre de Cournot-Nash / ***Cournot-Nash equilibrium***
Combinaison stratégique telle que la stratégie de chaque joueur est une meilleure réponse aux stratégies des autres joueurs.

Équilibre de Bertrand / ***Bertrand equilibrium***
Équilibre d'un marché où deux firmes produisant un même homogène se font concurrence en prix.

Équilibre général / ***General equilibrium***
Prix et quantités tels que tous les marchés de biens ou services soient simultanément à l'équilibre.

Équilibre partiel / ***Partial equilibrium***
On dit d'un marché qu'il est à l'équilibre (partiel) lorsque les quantités offertes et les quantités demandées sur ce marché sont égales.

Fonction de demande / ***Demand function***
Relation, normalement décroissante, entre prix et quantité demandée d'un bien ou service.

Fonction d'offre / ***Supply function***
Relation, normalement croissante, entre prix et quantité offerte d'un bien ou service.

Fonction d'utilité / ***Utility function***
Fonction numérique attachée à un consommateur et qui associe à tout panier de biens une valeur d'autant plus forte que le panier est apprécié par ce consommateur.

Input (ou facteur de production) / ***Input***
Ingrédient combiné avec d'autres dans le cas d'un processus de transformation conduisant à l'élaboration, la confection, la réalisation d'un bien ou service (nommé output).

Isoquant / ***Isoquant***
Ensemble des combinaisons de deux inputs pour lesquelles il est possible d'atteindre une certaine cible de production d'output.

Loisir / ***Leisure***
En économie, temps non consacré au travail salarié.

Long terme / *Long run*
Délai à l'issue duquel, dans le cadre du modèle de concurrence pure et parfaite, i) toutes les firmes qui produisent le bien ou service ont adopté la technologie de production la plus efficace disponible et ii) le nombre de firmes présentes s'est ajusté de manière à ce que le profit microéconomique de chacune d'entre elles soit nul.

Monopole / *Monopoly*
Secteur d'activité où une firme unique est l'offreur du bien ou service produit et mis en vente.

Monopole naturel / *Natural monopoly*
Situation dans laquelle il est économiquement justifié et rationnel qu'une firme unique produise la totalité des unités proposées à la vente.

Monopsone / *Monopsony market*
Une firme est en position de monopsone si elle est l'unique acheteur d'un de ses facteurs de production.

Oligopole / *Oligopoly*
Secteur d'activité dans lequel un petit nombre de firmes de taille substantielle produisent et mettent en vente un bien ou service. Dans un oligopole, les firmes s'identifient toutes et interagissent stratégiquement.

Optimum de Pareto / *Pareto optimum*
Une allocation est optimale au sens de Pareto si aucun individu ne peut accroître sa satisfaction sans détériorer celle d'un autre au moins À l'inverse, s'il est possible de trouver une manière quelconque d'accroître la satisfaction d'un ou plusieurs individus sans pénaliser quelqu'un d'autre, il s'agit d'une allocation non optimale ; s'il n'est pas possible de trouver une telle amélioration, l'allocation est optimale.

Output / *Out put*
Bien ou service produit à l'issue de la combinaison de multiples facteurs de productions (nommés inputs).

Panier de biens / *Basket (or bundle) of goods*
Collection de biens ou services que consomme un individu.

Pouvoir de marché / *Market power*
Mesure de la capacité d'une firme à réaliser un profit microéconomique élevé sur un marché.

Productivité marginale / *Marginal product*
Surcroît de production d'output induit part un accroissement infinitésimal de la quantité utilisée d'un seul input.

Profit microéconomique / *Profit*
Fonction qui mesure, pour toute quantité produite d'un bien ou service, la différence entre la Recette Totale tirée de la vente des unités de ce bien et le coût total de leur production.

Recette marginale / *Marginal revenue*
Surcroît de recette lié à la vente d'une unité supplémentaire d'un bien ou service.

Relation de préférence ou indifférence / *Weak order (Consumer preferences)*
Relation d'ordre qui incarne l'expression des préférences d'un consommateur entre toute paire de paniers de biens.

Rendements d'échelle / *Returns to scale*
Surcroît de production d'output induit par un accroissement infinitésimal, de même ampleur, de la quantité utilisée de tous les inputs simultanément.

Surplus d'un consommateur / *Consumer surplus*
Équivalent monétaire du bien-être éprouvé par un consommateur à l'occasion de la consommation d'un bien ou service. Le surplus est égal à l'écart entre la disponibilité maximale à payer du consommateur pour ce bien et le prix qu'il paye effectivement pour l'acquérir.

Taux marginal de substitution / *Marginal rate of substitution*
Mesure des proportions dans lesquelles un consommateur est prêt à échanger un bien contre un autre sans que ne soit modifié le niveau de sa satisfaction.

Taux marginal de substitution technique / *Marginal rate of technical substitution*
Mesure des proportions dans lesquelles il est possible de substituer un input par un autre sans que la quantité d'output produite ne soit modifiée.

Technologie de production / *Production function*
Processus de transformation basé sur la combinaison de multiples facteurs de productions (nommés inputs) et conduisant à l'élaboration, la confection, la réalisation d'un bien ou service (nommé output).

Termes de l'échange / *Terms of trade*
Proportions dans lesquelles se fait l'échange, entre deux agents, d'un bien contre un autre.

Théorie des jeux / *Game theory*
Discipline mathématique qui étudie les décisions des agents dans un contexte d'interactions stratégiques.

Utilité marginale / *Marginal utility*
Surcroît de satisfaction engendré par la consommation d'une unité supplémentaire d'un bien ou service.

Welfare* (ou surplus global) / *Welfare
Mesure du bien-être éprouvé par une société. Le *welfare* est égal à la somme des surplus des consommateurs et des producteurs, c'est-à-dire à la somme des surplus des consommateurs et des profits.

Bibliographie

Etner J. et Jeleva M., *Microéconomie*, Dunod, 2014

Gayant J.-P., *Risque et décision*, Vuibert, 2001

Jullien B. et Picard P., *Eléments de microéconomie – 2. Exercices et corrigés*, Montchrestien, 2011

Mas-Colell A., Whinston M. D. et Green J. R., *Microeconomic Theory*, Oxford University Press, 1995

Pepall L., Richards D.J. et Norman G., *Industrial Organization: Contemporary: Theory and Empirical Applications*, Wiley, 2008

Picard P., *Eléments de microéconomie – 1. Théorie et applications*, Montchrestien, 2011

Salanié B., *Théorie des contrats*, 2e éd., Economica, 2012

Shy O., *Industrial Organization, Theory and Applications*, MIT Press, 1996

Tallon J.-M., *Equilibre général – Une introduction*, Vuibert, 1998

Tirole J., *Théorie de l'organisation industrielle – Tomes 1 et 2.*, Economica, 1993, 1995

Varian H., *Analyse microéconomique*, De Boeck, 2008

Varian H., *Introduction à la microéconomie*, 7e éd., De Boeck, 2011

Wasmer E. *Principes de microéconomie – Méthodes empiriques et théories modernes*, Pearson, 2013

Index

D

E

F

I

L

M

O

P

R

S

T

U

V

W